Predictive biomarkers of breast ductal carcinoma *in situ* may underestimate the risk of recurrence due to *de novo* ipsilateral breast carcinoma development

Tanjina Kader[1,2*$], Maia Zethoven[1*], Sakshi Mahale[1*], Hugo Saunders[1], Lauren Tjoeka[1], Rebecca Lehmann[1], Madawa Jayawardane[1,2], Jia-Min Pang[1], Dorothea Lesche[1], Neeha Rajan[3], Timothy Semple[1], Jue Er Amanda Lee[1,2], Richard Lupat[1], David J Byrne[1], Siobhan Hughes[1], Hoa Nguyen[1], Siqi Lai[1$], Maree Pechlivanis[1], Olivia Craig[1,2], Lisa Devereux[1,2], Eloise House[4], Sureshni I Jayasinghe[4,5], Tom L Kaufmann[6,7,8], Roland F Schwarz[6,7], Andrew R Green[9], Islam Miligy[9,10], Margaret Cummings[11], Sunil Lakhani[11,12], Ian G Campbell[1,2], Emad Rakha[9,13], Stephen B Fox[1,2], G Bruce Mann[1,2,3], Kylie L Gorringe[1,2#].

[1]Peter MacCallum Cancer Centre, 305 Grattan St, Melbourne, Australia, 3000.

[2]The Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Australia, 3010.

[3]The Breast Service, The Royal Women's Hospital, Melbourne, Australia.

[4]Department of Anatomical Pathology, St Vincent's Hospital Melbourne, Fitzroy, Victoria, Australia.

[5]Department of Clinical Pathology, Melbourne Medical School, The University of Melbourne, Parkville, Victoria, Australia.

[6]Institute for Computational Cancer Biology (ICCB), Center for Integrated Oncology (CIO), Cancer Research Center Cologne Essen (CCCE), Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany.

[7]BIFOLD - Berlin Institute for the Foundations of Learning and Data, Berlin, Germany.

[8]Berlin Institute for Medical Systems Biology (BIMSB), Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany.

[9]Nottingham Breast Cancer Research Centre, Academic Unit of Translational Medical Sciences, School of Medicine, University of Nottingham and Department of Histopathology, Nottingham University Hospitals NHS Trust, City Hospital, Nottingham, UK.

[10]Pathology Department, Faculty of Medicine, Menoufia University, Egypt.

[11]Pathology Queensland, Royal Brisbane and Women's Hospital, Brisbane, QLD.

[12]Centre for Clinical Research, the University of Queensland, Brisbane, QLD.

[13]Pathology Department, Hamad Medical Corporation, Doha, Qatar.

Corresponding author: kylie.gorringe@petermac.org

[*] These authors contributed equally to this work

[$] TK: Current affiliation: Lab of Systems Pharmacology, Harvard Medical School, Boston, MA; SL: Current affiliation: Department of Genetics, The University of Texas, MD Anderson Cancer Center, Houston, TX, USA

The authors declare no potential conflicts of interest

Abstract:

Purpose: Development of ipsilateral breast carcinoma following a diagnosis of breast ductal carcinoma *in situ* (DCIS) has been assumed to represent recurrence of the primary tumour. However, this may not be the case and it is important to know how often recurrences are new primary tumours to ensure appropriate individualised therapy.

Experimental Design: Ipsilateral primary-recurrence pairs (n=78) were sequenced to test their clonal relatedness. Shared genetic events were identified from whole exome sequencing (n=54 pairs) using haplotype-specific copy number and phylogenetic analysis. The remaining pairs were sequenced by a targeted panel or low-coverage whole genome sequencing. We included 32 non-recurrent DCIS to compare the genetic profiles between recurrent and non-recurrent disease to develop a predictive biomarker.

Results: We found that 14% of DCIS recurrences were non-clonal, indicative of a new breast carcinoma. Four chromosomal changes (5q, 11q, 17q and 20q) and *TP53* mutation were enriched in clonal primaries compared with non-recurrent DCIS ($p<0.05$, Fisher's exact test). The prognostic value of TP53 was validated in an independent cohort using immunohistochemistry (HR=3.1; 95% CI 1.3-7.8). Non-clonal DCIS primaries had a very similar genetic profile to non-recurrent DCIS, suggesting this subset of cases would be identified as being at "low risk" of recurrence using tumour-intrinsic markers.

Conclusions: We have identified a substantial rate of new ipsilateral primary carcinomas after a diagnosis of DCIS. Our results suggest, as with invasive breast cancer, that if a recurrent tumour with an independent origin occurs, then the patient is at a high risk through the environment, the breast microenvironment and/or has a genetic predisposition. Importantly, the frequency of new primaries will influence the interpretation of findings in DCIS biomarker discovery studies as the true recurrence rate will be incorrect, affecting these efforts.

Translational relevance:

Our finding that >10% of recurrent tumours are new primaries provides genetic evidence that the presence of DCIS confers a risk of a *de novo* breast cancer as well as recurrence. Identifying a biomarker of such risks might allow preventive actions, such as genetic testing, chemoprevention with tamoxifen or aromatase inhibitors, or bilateral mastectomy. The corollary of these findings is that *de novo* primaries in DCIS biomarker studies may have undermined efforts to find a biomarker of recurrence by reducing statistical power, since a tumour cell-intrinsic marker is unlikely to be predictive for a new primary. Even if a tumour molecular biomarker could stratify between non-recurrent and recurrent patients, it will under-detect patients at risk of new primaries. This issue raises concerns about utilising only a tumour cell-intrinsic biomarker in the clinical setting.

Keywords: ductal carcinoma *in situ*, recurrence, clonality, breast neoplasm, whole exome sequencing, phylogenetic analysis

Introduction

Ductal carcinoma *in situ* (DCIS) is a pre-invasive breast lesion long known as a non-obligate precursor of invasive breast cancer (IBC). Because up to 25% of DCIS may recur as either DCIS or IBC (1), DCIS is treated with surgery with or without radiotherapy and/or systemic endocrine therapy. It has long been recognised that this approach overtreats many women due to the lack of accurate prognostic markers of recurrence (2, 3). Treatment options remain the same for recurrent patients with further surgery +/- radiotherapy for recurrences of DCIS, and additional systemic therapies if indicated for IBC.

There has been an assumption that all ipsilateral breast carcinomas that develop after a diagnosis of DCIS are genetically related (i.e. clonal) to the primary tumour (true recurrences), but this may not always be the case. Using somatic copy number alterations (CNA) derived from SNP arrays, we previously showed that 2/8 recurrences were non-clonal (4). Others have reported using limited resolution methodology that 18% of IBC recurrences are non-clonal (i.e. new primaries), and four cases were determined to be independent lineages to the primary DCIS using single cell DNA sequencing (5). Knowing the true recurrence rate is important not only to guide patient management but also critical in designing any study aiming to identify DCIS biomarkers of recurrence. If the rate of non-clonal recurrences is high, then tumour-intrinsic based biomarker discovery is compromised and strategies will be needed to identify those at risk of new primaries.

Here we assembled a large cohort to explore the clonal relatedness and evolutionary trajectory of primary DCIS and subsequent carcinoma pairs. For simplicity, we will refer to all such second events as “recurrences”, even when non-clonal. We performed genetic analysis of both non-recurrent DCIS and primary-recurrence tumour pairs with high depth

whole exome sequencing (WES), low-resolution whole genome sequencing or a targeted sequencing panel to provide an insight into the molecular features of recurrences.

Materials and Methods

Patient samples

Primary DCIS and recurrent DCIS/IBC (n=82, ipsilateral n=78 and contralateral n=4) as well as non-recurrent DCIS (n=32) cases were identified through hospital databases from Royal Melbourne Hospital (RMH), Nottingham City Hospital (UK), the prospective LifePool cohort and Peter MacCallum Cancer Centre (PMCC). Patients were recorded as non-recurrent when there was no record of a second incidence of DCIS/IBC within 7 years of the initial DCIS diagnoses (median follow-up was 8.5 years, range 7-23). We included patients as non-recurrent when they had undergone breast conserving surgery (i.e. no mastectomy) with a clear/negative margin. In contrast, recurrent patients were diagnosed with either DCIS or IBC >1 year following their initial DCIS diagnosis regardless of the surgery type. The cases are described in Supp File 1. Overall, 55% of ipsilateral recurrences (43/78) were IBC. Archival FFPE blocks of non-recurrent DCIS, all primary DCIS and recurrent tumours were obtained from the associated hospitals and haematoxylin and eosin stained slides were reviewed by pathologist(s) to choose the best areas for microdissection. Stromal regions from 10 breast samples were micro-dissected and used in each sequencing run as a pooled normal control and for validating rare variants. The summary of the cohorts, sequencing technologies and aims are summarised in Figure 1.

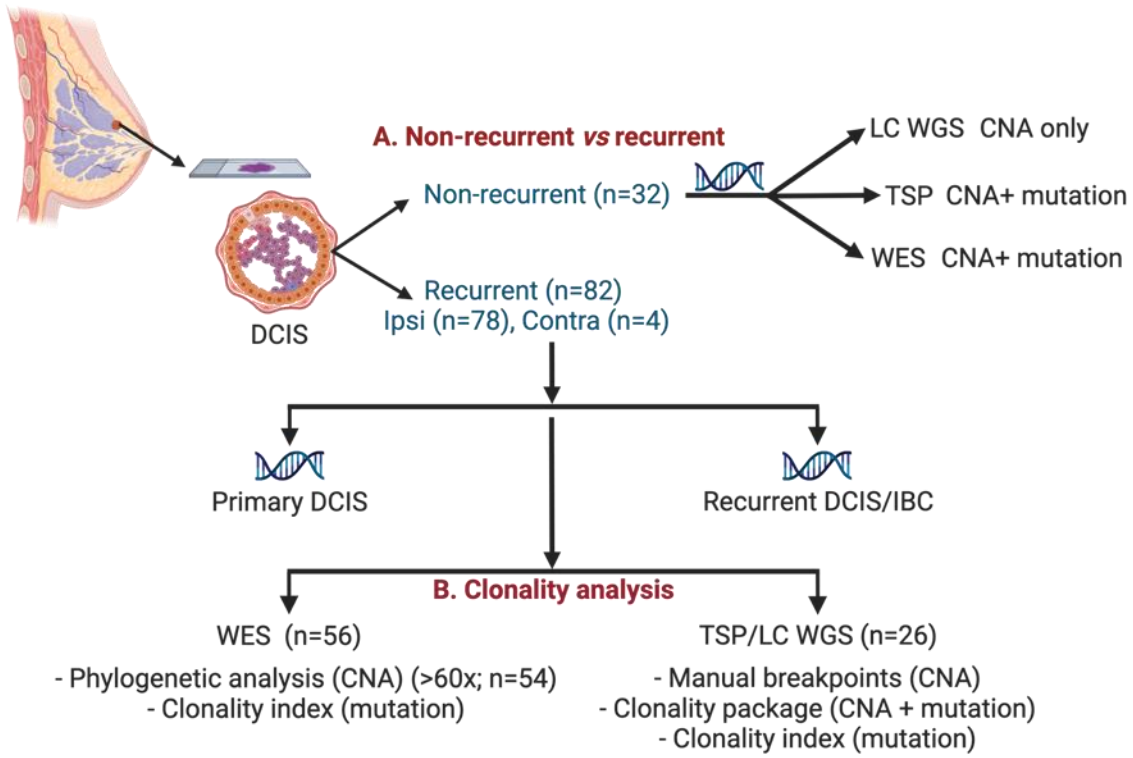


Figure 1. Experimental design. Non-recurrent DCIS (n=32) and 82 pairs of matched primary DCIS-recurrent tumours were microdissected followed by DNA extraction. Depending on the sequencing technology, clonality analyses were performed either by inspecting manually for copy number alterations (CNA) and the Clonality package or by investigating evolutionary history by phylogenetic analysis. Mutation profiles were generated either by targeted sequencing panel or WES and clonality analysis was performed by the Clonality package and clonality indices CI and CI2. TSP, targeted sequencing panel; LCWGS, low coverage whole genome sequencing; WES, whole exome sequencing.

This study was conducted under ethical approval from the Peter MacCallum Cancer Centre (HREC#16/PMCC/122), Melbourne Health (HREC#2008.219), St Vincent's Hospital (HREC#022-19) and the North West-Greater Manchester Central Research Ethics Committee 15/NW/0685.

Tissue micro-dissection and DNA extraction

All cases were micro-dissected from 7-10 micron haematoxylin stained sections (range 4-20 sections) by manual micro-dissection to achieve >50% tumour tissue purity. DNA was extracted using the FFPE AllPrep kit (Qiagen, Germantown, MD, USA) as per the

manufacturer's protocol. Only DNA was used in this study and Quant-iT™ dsDNA High-sensitivity Assay Kit (Invitrogen, Carlsbad, CA, USA) was used to determine DNA concentration. The quality of DNA was assessed by a multiplex PCR assay with primer sets that produce 100 - 700 bp fragments from the *GAPDH* gene as described previously (6). Depending on the success and timing of DNA extraction (Figure 1), cases were analysed by a targeted sequencing panel (n=46, Nottingham and LifePool samples pre-2020), whole exome sequencing (WES, n=67, RMH, PMCC and remaining LifePool cases), or low-coverage whole genome sequencing (LCWGS). LCWGS was chosen for the cases for which <30 ng DNA was available.

Only rare and uncommon pathogenic variants identified by WES were validated by Sanger sequencing with matched normal DNA extracted from stroma to rule out germline *vs* somatic variants as previously described (7).

Library preparation for WES

Paired primary-recurrent tumour DNA was processed by the Australian Genome Research Facility (AGRF) using the Twist Bioscience Human Comprehensive Exome v1 or v2 (Twist Bioscience HQ, South San Francisco, CA, USA) according to the manufacturer's protocol (Twist Target Enrichment Standard Hybridisation Protocol). Sequencing was performed on the Illumina NovaSeq 6000 platform with 150 bp paired-end reads for a median depth of 101.29x, range: 20.39-247x. Each run had at least one stromal DNA sample as a control.

Targeted sequencing library preparation, enrichment and sequencing

Targeted sequencing of tumour DNA was performed using an Agilent SureSelect Custom Panel targeting known IBC driver genes from large cohort studies, such as the Cancer

Genome Atlas (TCGA). This panel has previously been used (7, 8) and tailored for IBC and DCIS driver genes such as *PIK3CA, AKT1, PIK3R1, GATA3, PTEN, TP53, ARID1A* (9) (gene list: Supplementary File 2). Library preparation was performed by the Molecular Genomics Core, PMCC, mostly from an input of at least 100 ng of DNA using the KAPA HyperPrep Kit (Roche Diagnostics, Rotkreuz, Switzerland) and the SureSelect XT (Agilent, Santa Clara, CA, USA) hybridisation capture system as described previously (7, 8). Sequencing of target-enriched DNA libraries was performed using the Illumina Next Seq 500 generating 75 bp paired-end sequence reads. Each run had at least one stromal DNA sample as a control.

DNA Library Prep and LCWGS

A low DNA input library preparation protocol was used only for samples with <30 ng DNA, using the NEBNext® Ultra ™ II DNA Library Prep Kit (NEB E7645S/L, New England BioLabs ® Inc., Ipswich, MA, USA) as described previously (6, 7, 10). These libraries were used for LCWGS as described (6, 7, 10). Briefly, an Illumina Nextseq platform (NextSeq 500) (Illumina, San Diego, CA, USA) (paired-end 75 bp) was used to run the pooled, normalised indexed libraries according to the standard Illumina protocol. The sequencing depth achieved in the samples was approximately 2x.

Whole Genome Sequencing (WGS)

Libraries for four paired samples with matched normal (stromal DNA) were made, processed and sequenced by AGRF using the IDT xGen Kit (USA) according to the manufacturer's protocol, aiming to achieve 60x depth for tumour DNA and 30x for matched normal.

Data analysis for WES and targeted sequencing panel

For targeted sequencing panel and WES, paired-end sequence reads were aligned to the g1k v37 hg19 reference genome using BWA (11). Optical duplicate reads were removed using Picard (http://broadinstitute.github.io/picard/) (v1.119), then local realignment around indels and base quality score recalibration were performed using the Genome Analysis Tool Kit (GATK v3.8) in accordance with their recommended best practice workflow (12). SNP and indel variants were called using GATK Unified Genotyper, Platypus (13) and Varscan 2 (14). Called variants were additionally annotated using the Ensembl Variant Effect Predictor release 78 (15).

Estimating somatic mutations from targeted panel and WES

Somatic mutations in the tumour sequencing data were identified by applying the following filters: canonical transcript; variants identified by at least two variant callers, allele depth >5 and read depth >10 unless the variant was in *PIK3CA* or *TP53*. Since matched normal DNA was unavailable for this study, any variant with minor allele frequency present in the GnomAD non-cancer population (version 2.0) unless known to be pathogenic (e.g. *TP53*) was filtered out. Such pathogenic variants were validated along with their matched normal DNA by Sanger sequencing as described previously (16) to ensure their classification as a somatic mutation prior to clonality analysis. Any rare germline variant misclassified as somatic mutation could lead to a false clonality signal (17). Therefore, this step of filtering out from GnomAD non-cancer population was strictly maintained (18). Variants reported in the literature as sequence artefacts were filtered out (19). Manual inspection of the sequence reads using the Integrative Genomics Viewer (20) was performed before finalising the somatic mutations. Any false positive variants due to sequencing artefacts were excluded from the final analysis.

Estimating CNA from targeted panel and WES and multi-sample phasing from WES

Both off-target and on-target sequencing reads from the targeted panel were used to generate genome-wide copy number data using PureCN without matched normal (21). However, 24 normal DNA samples were pooled and used as a baseline of PureCN for most WES cases (Exome v1) except for 12 primary-recurrent pairs for which only 1 normal DNA was used (Exome v2). Out of multiple PureCN solutions/case for CNA profiles, purity and ploidy, the first solution (solution 1) was used for downstream analysis.

CNA profiles, purity and ploidy status from solution 1 of PureCN data from WES was then used to generate haplotype-specific copy number changes by multi-sample phasing using Refphase (an R package for reference phasing copy number segments) as described by Watkins *et al.* (22) (https://bitbucket.org/schwarzlab/refphase/src/master/). Samples with insufficient depth and low quality failed to phase and were excluded (n=2: Case DCIS00191, DCIS00401). Haplotype-specific copy number changes by Refphase revealed mirrored subclonal allelic imbalance (MSAI) (i.e. parallel evolutionary events) and were leveraged to infer phylogeny reconstruction and ancestral genomes between primary-recurrent pair by Minimum Event Distance for Intra-tumour Copy-number Comparisons-2 (MEDICC2) (n=54/56) described below.

Estimating CNA from LCWGS

For LCWGS, reads were aligned with bwa mem (v0.7.12-r1039) to hg19 (GRCh37) after removal of sequencing primers by cutadapt (v1.7.1) as described previously (6, 7, 10, 23) . ControlFREEC (version 6.7) (24) was used to estimate copy number from the LCWGS data in 50 kb windows, with default parameters, no matched normal sample and baseline ploidy set to 2. To reduce spurious calls, blacklisted regions as identified from Scheinin *et al*. (19)

were excluded. In addition to the CNA, fraction of genome altered (FGA) was calculated as previously described (4).

Clonality analysis

Manual inspection of copy number breakpoints: All sample $\log_2$ ratio raw data generated by PureCN were imported into Nexus (v10, BioDiscovery Inc., Hawthorne, CA) and segmented using SNP-FASST with a stringency of $1x10^{-8}$ for WES and targeted sequencing panel and $1x10^{-5}$ for LCWGS samples. Copy number gains were called if the $\log_2$ ratio of the segment was >0.18 and losses were called if <-0.18 as previously described (4, 7, 10). Manual inspection of break points was performed to define the genetic relatedness of a primary-recurrent pair (4, 7, 10, 17, 25). Paired cases sharing a non-centromeric breakpoint or >2 centromeric breakpoints with the same CNA were defined as clonal, with the exception that if these shared events were only common breast cancer CNA (1q gain, 16q loss or 8q gain) then clonality was defined as equivocal.

Clonality package: In addition to the manual inspection, a statistically based clonality analysis, Clonality, an R package (v 3.6), was used based on total CNA and mutations (26, 27). This statistical approach was only used for cases run by targeted sequencing panel for CNA. All samples were run together in the package to generate the reference distribution (27). The paired cases were only analysed when both primary and recurrent pairs were sequenced with the same technology. Since the package was developed based on low resolution comparative genomic hybridisation and considers just one prominent CNA per chromosome arm, targeted sequencing panel data were reduced to ~10,000 locations as suggested. A *p*-value of 0.01-0.05 was considered as equivocal, $p<0.01$ was clonal and $p>0.05$ was non-clonal (Supp Figure 1). Paired genome-wide copy number plots for all cases

were generated as described in the package and they all were manually verified. This manual verification was necessary for all cases, especially if there was any discordance between manual breakpoint inspection and the Clonality package. FFPE samples are notorious for spurious calls, therefore, any apparent rare or uncommon very small segment of gain/deletion might generate a low *p*-value and be considered clonal by the package (Supp Figure 2). Manual verification was also needed to rule out any sequencing-specific artefacts, such as 19p which often is seen as gained on the targeted sequencing panel (Supp Figure 3). For somatic mutation, both the targeted sequencing panel and WES data were considered. A *p*-value generated by the Clonality package "get.mutation.frequencies" function suggests a clonal pair if $p<0.05$, with TCGA breast cancer cohort used as a reference cohort as suggested by Mauguen *et al.* (26). The *p* values generated by this package for both CNA and mutational profiles are recorded in Supp File 3.

<u>Phylogeny reconstruction using MEDICC2:</u> An evolutionary method was used for most paired cases in this study with high-depth WES data (n=54/56) and WGS data (n=4). Cases with <20x depth or low quality were excluded (n=2: Case DCIS00191, DCIS00401). Copy number profiles were pre-phased using Refphase to derive haplotype-specific CNA profiles (22) and MSAI, which then was used to derive phylogenetic trees for tumour pairs using MEDICC2 (28). An example of Refphase profile is shown in Supp Figure 4. All Refphase and MEDICC2 profiles were manually inspected to reduce overcalling clonality due to the presence of small shared segments, which were likely FFPE related artefacts.

<u>Clonality index only for somatic mutations:</u> An additional estimation for clonal relatedness was performed described by Schultheis *et al*. (18), called Clonality Index (CI) and CI2. This

approach was carried out for all cases run by targeted sequencing panel and WES. The CI for a tumour pair was defined as:

$$CI = \begin{cases} 1 - \prod_{k=1}^{n} f_k \, , n > 0 \\ 0, n = 0 \end{cases}$$

where f_k is the percentage of tumours in TCGA (29) harbouring a given mutation (k) and n is the number of shared mutations between a pair of tumours. The pair is considered genetically related if CI is more than 0.8.

CI2 is similar to CI, but by using an R package ROCR. Unlike CI, where the threshold was fixed as 0.8 to define any pair tumour as "clonal", the CI2 threshold will be carried out by the ROCR package. The threshold will vary based on the average mutation rate of that tumour type in the control dataset (TCGA-BRCA) as well as the frequency distribution of variants of the testing cohort. In our cohort, the generated threshold by the R package was 2.83, indicating any pair with CI2 value >2.83 was considered as clonal. Both CI and CI2 values are detailed in Supp File 3.

Histology and Tumour Infiltrating Lymphocytes (TILs)

Estrogen receptor (ER) was determined as reported on the pathology record at the time of diagnosis or was performed as described (30). HER2 amplification status was determined from copy number data (4). TILs of non-recurrent and recurrent cases were available from pathological review at the time of diagnosis or assessed later for this study by breast pathologists (J.M.P. or N.R.). For recurrent cases, TILs were assessed from both primary DCIS and recurrent tumours. The method for assessing TILs was developed by Pruneri *et al.* (31) for DCIS and supported by the recent guidelines of the International Immuno-Oncology Biomarker Working group (32-34).

TP53 immunohistochemistry

Tissue microarrays were obtained from Nottingham, UK (30) and Queensland, Australia. TP53 IHC was performed by the Peter MacCallum Dept of Anatomical Pathology using the standard clinical assay (clone DO7) as previously (16). Scoring was performed blinded to mutation status and outcome, and required at least 50 tumour cells to be present in the core. Cases were called normal if there was heterogeneous staining in <50% of cells and no contiguous areas of staining. Cases were abnormal if they carried one of three mutation-related patterns (35). Overexpression (OE) included strong nuclear staining in >5% contiguous cells and required contiguously stained cells across at least 50% of a duct containing at least 20 cells. Complete absence (CA) was no staining in any tumour cells, but positive staining in some stromal or lymphocyte cells. Cytosolic staining (CY) was diffuse cytoplasmic staining in cells in the same proportions as for OE. CY may or may not have nuclear staining also.

Statistical analysis and data visualisation

GraphPad Prism v 9.2.0 (San Diego, CA, USA), R v.4.3.2 (36) and RStudio (v 2023.09.1, Posit Software) were used to generate graphs, perform statistical tests and analyses. A *p*-value of <0.05 was considered significant unless stated otherwise. Tables and associated statistical tests results were generated using *gtsummary* (v 1.7.2)(37). The default test for continuous variables was the Wilcoxon test, for contingency tables chi-squared unless expected values were <5, in which case a Fisher's exact test was performed. Univariate and multivariable Cox regression were performed to identify clinical and molecular characteristics associated with patients' recurrence risk (*survival::coxph* v 3.5-7) (38). All cases with data for each characteristic were used in univariate analyses (n=186), however only the subset with complete data due to missing values was used in a multivariate analysis

(n=134). The proportionality assumption was checked using *cox.zph*. Internal model calibration and validity tests were performed (*rms* package v 6.7-1(39)). Adjusted and unadjusted survival curves were generated using the *survminer* package v 0.4.9 (*ggadjustedcurves* and *ggsurvplot*)(40).

Figures were produced using Biorender or R and assembled in Adobe Illustrator. Copy number frequency plots were created using *GenVisR:cnFreq* (41). Oncoprint image was produced through the cBioPortal (42)

Data Availability

The sequencing datasets supporting the conclusions of this article will be available through NCBI Sequence Read Archive.

Results

Clinicopathological characteristics of the cohort

We performed genomic analysis on 82 pairs of DCIS with recurrence and 32 DCIS without recurrence. Out of 82 pairs, four pairs were contralateral and anticipated to be non-clonal controls but not included in summarised information or comparisons. When we compared the recurrent and non-recurrent cohorts, there was no significant difference in patients' age at the initial diagnosis of DCIS, ER/HER2 status of primary DCIS, or the size of primary DCIS (Table 1). Recurrent cases were more likely to be high grade (p=0.047) and to have been treated without radiotherapy (p<0.001) (Table 1, with further detail in Supp File 1).

Table 1. Cohort characteristics

Variable	**Non-recurrent**, N = 32[1]	**Recurrent**, N = 78[1]	**p-value**[2]
Age (years)	61.3 (±7.5)	60.5 (±9.5)	0.6
Unknown	0	3	
Radiotherapy	11 / 17 (65%)	13 / 71 (18%)	<0.001
Unknown	15	7	
Grade			0.047
High	14 / 32 (44%)	51 / 78 (65%)	
Intermediate	14 / 32 (44%)	16 / 78 (21%)	
Low	4 / 32 (13%)	11 / 78 (14%)	
Size (mm)	15.1 (±12.2)	18.8 (±17.1)	0.6
Unknown	0	2	
ER			0.2
Negative	4 / 32 (12.5%)	17 / 75 (23%)	
Positive	28 / 32 (87.5%)	58 / 75 (77%)	
Unknown	0	3	
HER2			0.6
Negative	24 / 32 (75%)	55 / 78 (71%)	
Positive	8 / 32 (25%)	23 / 78 (29%)	
Tumour lymphocytes	11.3 (±8.4)	14.5 (±17.3)	0.8
Unknown	5	26	

[1]Mean (SD); n / N (%)

[2]Wilcoxon rank sum test; Fisher's exact test; Pearson's Chi-squared test

All but two of the recurrent tumours had the same ER status as the primary DCIS when data were available for both of the pair (Supp File 1). These two ER positive primary DCIS (LP004, DCIS00498) recurred as ER negative disease. Keeping in mind that IBC grade is not directly comparable to DCIS grade, the grade of recurrent tumours was not always similar to the primary DCIS. When grades of both were available (75/78 pairs), 23 out of 75 pairs (31%), recurred as a lower grade than primary (i.e. high or intermediate grade DCIS recurred as low grade DCIS or grade 1 IBC, e.g. DCIS00009 high grade DCIS recurred as low grade DCIS). On the other hand, 13 out of 75 pairs (18%), recurred as a higher grade than primary (for example LP003 low grade DCIS recurred as grade 2 IBC).

Evidence for *de novo* development of ipsilateral recurrent tumours

We first investigated 26 tumour pairs (4 contralateral) using a targeted sequencing panel (n=20) or LCWGS (n=1) or a combination (n=5) due limited DNA availability to test their clonal relatedness (examples in Supp Fig 1). Therefore, while CNA profiles were available for all pairs, mutation analysis was only available for 20 pairs. Clonal relatedness was assessed for this cohort by the Clonality package, manual inspection of breakpoints and clonality index (CI, CI2) (when mutation data was available).

Using copy number data, all contralateral tumours were genetically independent (i.e. non-clonal), as expected, while 68% (15/22) of ipsilateral primary-recurrent pairs were genetically related (i.e. clonal, representing true recurrences) and the remaining seven cases appeared to be independent tumours (i.e. non-clonal). Six of these seven cases had paired mutation data and all six pairs remained non-clonal when including these data. Out of these six cases, four pairs had no shared mutations or CNA, and all of these carried private CNA or mutations. Two out of these four pairs had private mutations in both primary and recurrent tumours, further evidence for the second tumours to be new primaries. One pair (Case 3933) had two shared copy number events but with different breakpoints including 16q loss, one of the most common copy number events of IBC (Clonality package p =0.98) (Supp Fig 3). The remaining pair (Case 3936) had no shared CNA but shared a *PIK3CA* (H1047R) mutation, the most common mutation of IBC. This result suggested a non-clonal pair by Clonality package and CI index (p=0.074, CI=0.8, CI2=2.83) (Supp File 3).

Phylogenetic analysis using haplotype-specific copy number profiles and somatic mutations confirmed independent ipsilateral recurrent tumours

The above preliminary data was based on relatively low-resolution technologies, resulting in limited power for clonality detection, such as manual inspection of CNA and breakpoints.

Manual inspection of breakpoints can confirm clonality but there can be uncertainty that cases with common CN events are related or not when lacking allelic information. Therefore, we attempted to obtain high depth WES data for the remaining paired primary-recurrent cases (n=56). Unfortunately, two cases were unable to achieve high sequencing depth for both samples (DCIS00191, DCIS00401). Clonality analysis was performed for these two cases by manual inspection of copy number breakpoints and clonality index as above (Supp File 3), which suggested clonal cases.

For the remaining cases, we undertook phylogenetic analysis using Refphase and MEDICC2. Refphase incorporates allelic information when assessing CN events. MEDICC2 is based on the minimum event distance between all genomes using neighbour joining taking into account whole genome doubling (WGD) events (43). As opposed to looking at any chromosomal event as a single event, such an analysis may reveal the evolutionary history of CNA. Supp Fig 5 shows an example of an *apparent* non-clonal pair as defined by manual inspection of shared CNA with different breakpoints, but that was classified as clonal (i.e. shared ancestral genomes) by investigating their evolutionary history.

Of the 54 pairs for which the sequencing depth enabled phylogenetic analysis, 49/54 (91%) shared an ancestral genome (i.e. clonal, Figure 2). When we investigated somatic mutations (Figure 3) for all of these cases, five out of six non-clonal pairs remained non-clonal due to lack of shared somatic mutations or sharing only *PIK3CA* H1047R (Case DCIS00506: CI=0.8, CI2=2.83, p=0.06 by Clonality package) (Supp File 3). Out of these five independent pairs, four of them with enough DNA were also subjected to whole genome sequencing with their matched normal stromal DNA to check their clonal relatedness. The CNA profiles and mutation status reconfirmed that three were non-clonal (e.g. Supp Figure 6), however,

DCIS00506 contained a small deletion at the end of chromosome 16 that, along with the shared *PIK3CA* mutation and 1q gain, suggested a clonal origin (Supp Figure 7). Taken together, high depth sequencing confirmed 4/54 cases as non-clonal (7.4%).

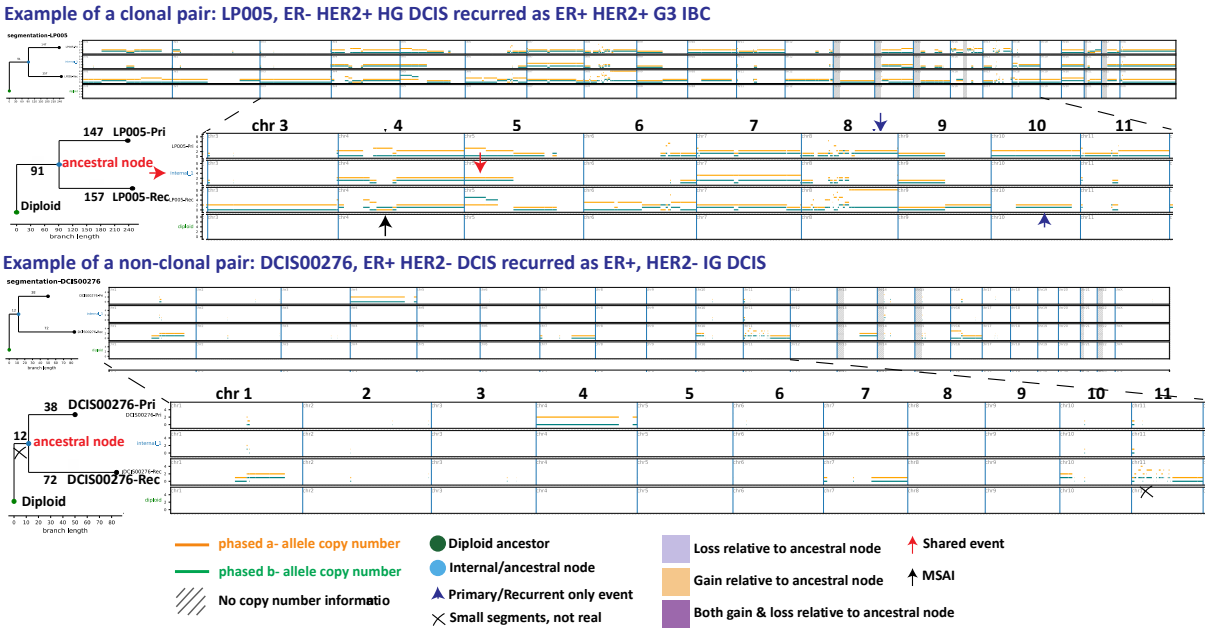


Figure 2. Example of a clonal pair and a non-clonal pair sequenced by high depth whole exome sequencing (WES). These profiles were generated by MEDICC2 based on haplotype-specific copy number profiles. Firstly, in case LP005 the phylogenetic tree suggests there were 91 clonal segments identified in the ancestral genome (internal node), indicating truncal events in the primary DCIS and recurrent IBC. The recurrent tumour had 157 divergent CNA segments and the primary had 147, indicating subclonal events for both tumours that occurred after the clonal events (examples: purple arrows). Chromosomes 3 to 18 magnified for visualisation. For example, gains of an "A-allele" of the first three segments of chromosome 5 ("A" allele had 2 copies while "B" allele had 1) were shared between primary and recurrent. Then there was a further gain of the A-allele (from 2 copies to 3 copies) and loss of heterozygosity of the B-allele (1 copy to 0) for the first 4 segments in the primary DCIS. On the other hand, the first two segments of chromosome 5 in recurrent IBC had gain of the B-allele to 4 copies and the first segment gained one more time for a total of 5 copies while the A-allele remained at 2. An example of mirrored subclonal allele imbalance (MSAI) is shown on chromosome 4 (black arrow) suggesting loss of different alleles on the same CNA segments (i.e. parallel evolution). For the case DCIS00276, a few small segments were detected as truncal events (truncal segments =12). Manual inspection of the Refphase profile and segmentation files derived from MEDICC2 suggested that they are not real segments, a well-known characteristic of FFPE samples. Lack of shared CNA and mutations suggested that this pair was non-clonal, which was confirmed by whole genome sequencing (Supp Figure 6).

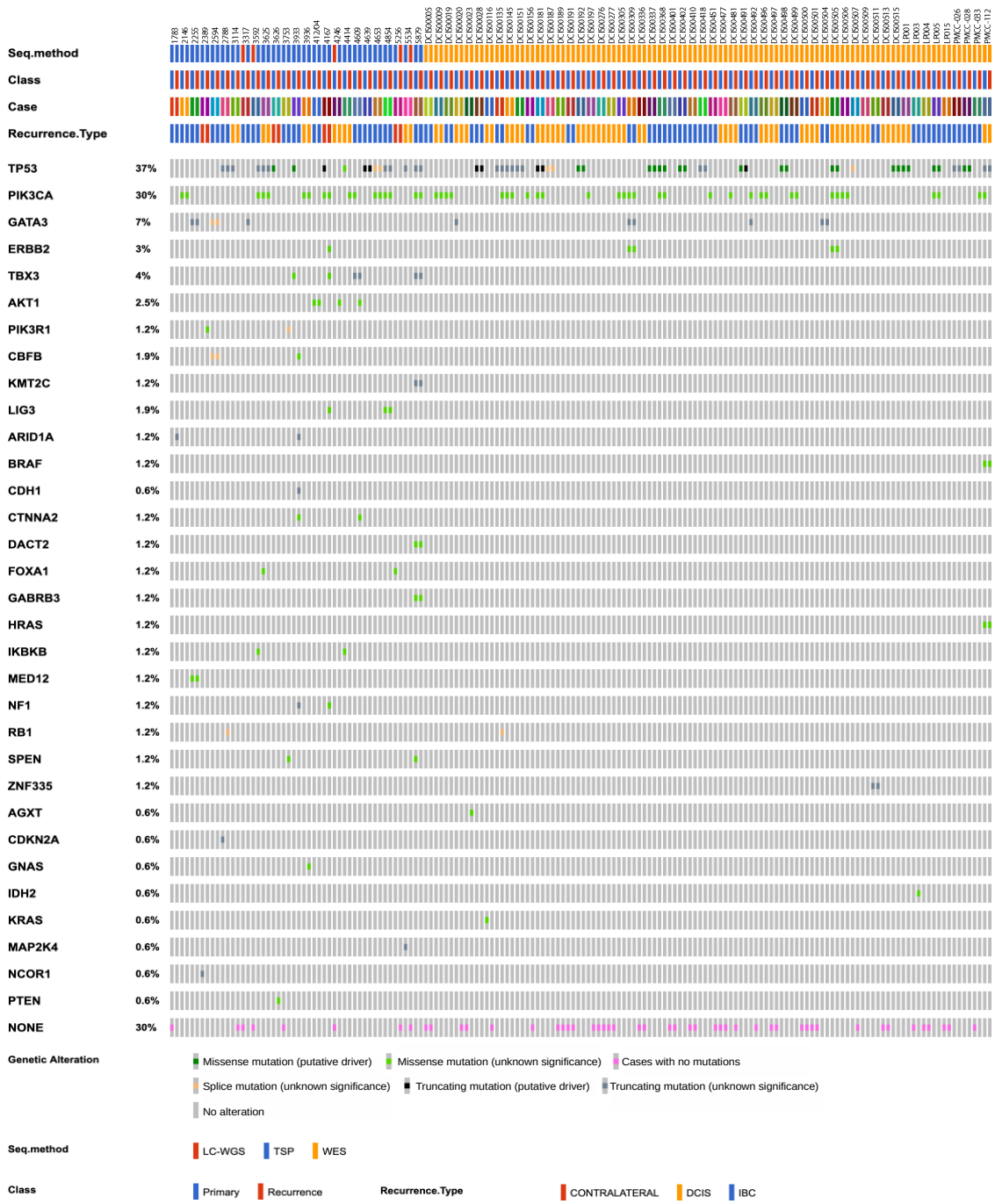


Figure 3. Oncoprint profile for somatic mutations of recurrent cases. For each paired case, both primary (blue) and recurrent (red) tumours are shown. Different types of mutation are indicated for each gene with the frequency of total mutation observed in a particular gene given. This profile was generated by the cBioportal visualisation tool(42).

The rate of detecting non-clonality was significantly different between high-resolution analysis by WES/WGS (4/54, 7.4%) compared to low-resolution by targeted

panel/LCWGS/low WES (7/22, 31.8%), p = 0.01 Fisher's exact test). Clonality detection was not strongly affected by purity of the tumour DNA, which was not significantly different between non-clonal and clonal primaries (t-test p=0.33) or recurrences (t-test p=0.98).

Combining all data, 11/78 cases (14%) were non-clonal, suggesting that these so-called recurrent tumours after the initial DCIS diagnosis are new primaries, including IBC recurrence (6/11 (55%), which is 7.7% of all recurrences).

Association of clonality status with clinico-pathological features

The clonality status (clonal recurrence *vs* non-clonal recurrence) was not significantly associated with the type of recurrence (IBC/DCIS, p=0.8), or the grade (p=0.6), ER status (p=0.7) or HER2 status (p=0.5) of the primary DCIS (Table 2, Supp Figure 8). Treatment with radiotherapy was also not associated with clonal recurrence (p=0.7). The age at diagnosis did not affect the type of recurrence (clonal median 60.0, non-clonal 59, p=0.8). Only the size of primary DCIS was significantly different between non-clonal (median 7.6 mm, range 2-17 mm) and clonal DCIS (median 15.5 mm, range 1.8-80 mm, p=0.023), but was not statistically significantly different to non-recurrent DCIS (median 15.0 mm, range 2-48 mm) (Supp Fig 8A). In addition, there was no significant difference in TILs among clonal-primary (median 10%, range 0-90%), non-clonal primary (median 5%, range 3-5%) and non-recurrent (median 5%, range 1-30%) (p=0.32, Kruskal-Wallis rank sum test) (Supp Fig 8B).

Table 2. Differences by recurrence clonality status

	Non-recurrent N = 32[1]	Clonal recurrence N = 67[1]	Non-clonal recurrence N = 11[1]	p-value[2]
Age	64.0 (43.0-73.0)	60.0 (42.0-84.0)	59.0 (31.0-78.0)	0.8
Unknown	0	2	1	
Radiotherapy	11 / 17 (65%)	12 / 62 (19%)	1 / 10 (10%)	<0.001
Unknown	15	5	1	
Grade				0.12
High	14 / 32 (44%)	45 / 67 (67%)	6 / 11 (55%)	
Intermediate	14 / 32 (44%)	13 / 67 (19%)	3 / 11 (27%)	
Low	4 / 32 (12.5%)	9 / 67 (13%)	2 / 11 (18%)	
Time to recurrence	N/A	3.3 (1.0-14.3)	4.7 (1.3-9.1)	
Type of recurrence				>0.9
DCIS	0 / 0 (NA%)	30 / 67 (45%)	5 / 11 (45%)	
IBC	0 / 0 (NA%)	37 / 67 (55%)	6 / 11 (55%)	
Unknown	32	0	0	
Size (mm)	15.0 (2.0-48.0)	15.5 (1.8-80.0)	7.6 (2.0-17.0)	0.072
Unknown	0	1	1	
ER status				0.4
Negative	4 / 32 (13%)	16 / 66 (24%)	1 / 9 (11%)	
Positive	28 / 32 (88%)	50 / 66 (76%)	8 / 9 (89%)	
Unknown	0	1	2	
HER2 status				0.7
negative	24 / 32 (75%)	46 / 67 (69%)	9 / 11 (82%)	
positive	8 / 32 (25%)	21 / 67 (31%)	2 / 11 (18%)	
TP53 mutation				0.026
MUT	4 / 29 (14%)	27 / 64 (42%)	3 / 10 (30%)	
WT	25 / 29 (86%)	37 / 64 (58%)	7 / 10 (70%)	
Unknown	3	3	1	
PIK3CA mutation				0.11
MUT	15 / 29 (52%)	20 / 67 (30%)	3 / 10 (30%)	
WT	14 / 29 (48%)	67 / 67 (70%)	7 / 10 (70%)	
Unknown	3	0	1	
FGA	0.2 (0.0-0.5)	0.2 (0.0-0.6)	0.0 (0.0-0.6)	0.021
Ploidy	2.0 (1.9-3.4)	2.1 (1.9-4.1)	2.1 (2.0-4.0)	0.11
Unknown	3	3	1	
WGD				0.7
N	10 / 10 (100%)	46 / 52 (88%)	4 / 4 (100%)	
Y	0 / 10 (0%)	6 / 52 (12%)	0 / 4 (0%)	
Unknown	22	15	7	
Tumour lymphocytes	5.0 (1.0-30.0)	10.0 (0.0-90.0)	5.0 (3.0-5.0)	0.3

	Non-recurrent N = 32[1]	Clonal recurrence N = 67[1]	Non-clonal recurrence N = 11[1]	p-value[2]
Unknown	5	18	8	

[1]Median (Minimum-Maximum); n / N (%);[2]Kruskal-Wallis rank sum test; Fisher's exact test. FGA, fraction genome altered. WGD, whole genome duplication.

One hypothesis we had was that later recurrences would be more likely to be independent tumours. However, the median time to detecting the second tumour was not significantly different: 3.3 years for clonal and 4.7 years for non-clonal recurrences, respectively (p=0.6, Wilcoxon rank sum test, Figure 4). Indeed, there were 8 unequivocally clonal recurrences occurring more than 10 years after the initial diagnosis, 5 IBC and 3 DCIS. All shared at least 2 cancer driving mutations and/or copy number breakpoints and also had divergent events such as additional variants or copy number changes. There was a trend for invasive recurrences to have a longer time to recurrence (median 4.5 years compared to 2.8 years for DCIS recurrences, p=0.05, Wilcoxon rank sum test, Supp Figure 9).

When looking at the consistency of clinico-pathological features between paired primary and recurrence depending on clonality status, there was no significant difference in the TILs between primary and clonal/non-clonal recurrences (p>0.05 for both, paired t-test). There was also no difference in whether grade was consistent – when considering just the low-high mismatches, only four were detected, all between clonal primaries and their recurrences. A stricter application of grade consistency (e.g. low to intermediate called a mismatch) found that 5/11 non-clonal primaries matched their recurrences, and 35/64 clonal (p=0.75, two-tailed Fisher’s exact test). ER status was also consistent regardless of clonality with 9/9 non-clonal recurrences and 54/56 clonal recurrences showing the same ER status. Therefore, neither grade nor ER status are good indicators of clonality.

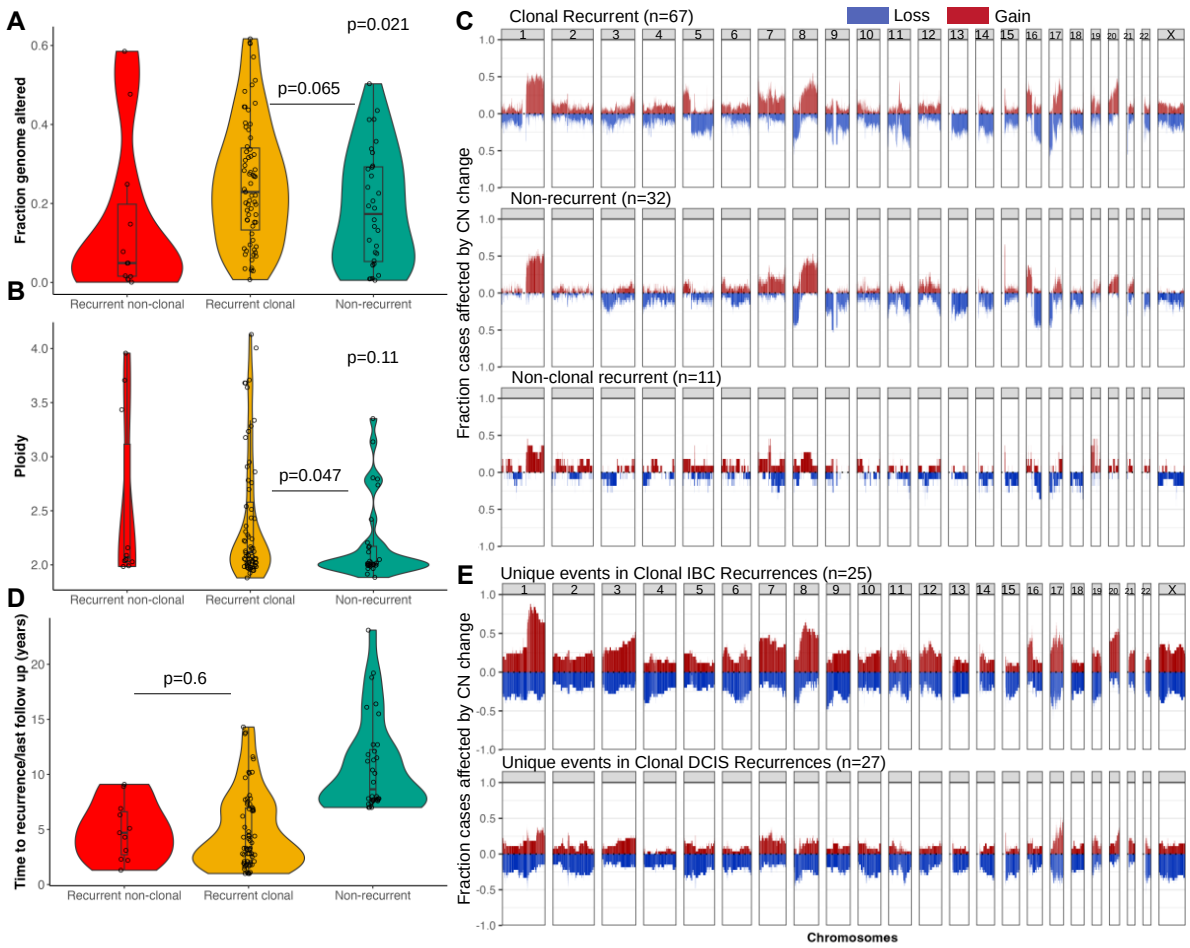


Figure 4. Comparison between non-recurrent and recurrent cohort. A. Fraction of Genome altered by copy number. **B**. Ploidy. The *P*-value at top right is the three-way comparison using Kruskal-Wallis test, while the value in the middle of the plot is the comparison between cases with a clonal recurrence and non-recurrent cases by a Wilcoxon rank sum test. **C**. Genome-wide copy number frequency of non-recurrent (n=32), cases with clonal recurrence (n=67) and non-clonal recurrence (n=11). Top=chromosome number (1-X), Blue=gain, Red=loss. **D.** Time to recurrence or last follow-up. Non-recurrent cases had at least 7 years follow-up. P-value is the comparison between cases with a clonal recurrence and non-clonal cases by a Wilcoxon t-test. **E.** Copy number frequency plot for regions specifically altered in the recurrence tumours (derived from WES samples analysed by MEDICC2). This includes regions where a CNA was already present in the truncal clone, and has been further changed in the recurrence sample. Note that when different alleles have been affected by different copy number events (e.g. gain of allele "A" and loss of allele "B"), both will be represented in the frequency plot.

Phylogenetic analysis showed the presence of ongoing chromosomal instability in DCIS and IBC recurrences

We next investigated the evolutionary history of the clonal pairs. Overall, the proportion of the genome affected by CNA was higher in the clonal recurrences (median 30.8%) than the index lesion (median 22.9%, p=0.027, paired t-test). This was primarily driven by an increase in FGA in the IBC recurrences (median increase of 13%, paired t-test *p*=0.0001). In contrast, FGA of non-clonal recurrent cases showed no significant difference between the matched pairs (median FGA 5% *vs* 10.6%, respectively) (p=0.89, paired t-test) (Supp Figure 9). However, MEDICC2 analyses showed that all clonal primary DCIS had the presence of subclonal events (range 30%-80%, median 55.7%), indicating ongoing chromosomal instability (Supp Figure 9E). There was no significant difference in terms of proportion of clonal CNA of primaries whether they recurred as DCIS or IBC (p=0.49, Wilcoxon test, Supp Figure 9F). The clonal diversification did not seem to be dependent on grade, ER or HER2 status or *TP53* mutation status of the primary DCIS (p all >0.05, Wilcoxon or Kruskal-Wallis test).

MEDICC2 revealed the presence of genome doubling events. No non-recurrent DCIS (0/10) or non-clonal de novo primaries (0/4) had WGD but 6/52 clonal primary DCIS (11%) and 8 clonal recurrences (15%) showed WGD events (p=0.7, Fisher's exact test). Genome-doubling events were truncal for three pairs, all with IBC recurrence (Supp Figure 10). All cases also had multiple subsequent gains and losses unique to primary and recurrence. When considering ploidy as an alternative measure of aneuploidy, overall there was no significant difference between primary DCIS and their clonal recurrences (p=0.23, paired t-test), but invasive recurrences had higher ploidy than their primary DCIS (p=0.03, paired t-test), similar to what we observed with the FGA (Supp Figure 8).

We further investigated CNA development in invasive recurrences to see whether any specific events were enriched. Using the segments flagged as being unique to the recurrence tumours by MEDICC2, we compared the frequency of new events in DCIS and invasive recurrences. As expected, the invasive recurrences had globally higher levels of CNA, but with notable increases in gain of 1q, 7q, 8q, 11q, 16p, and 20q (Figure 4).

Clonal diversification was also seen at the mutation level. Thirty-seven clonally related cases shared somatic driver mutations, most commonly in *TP53* (n=24) and *PIK3CA* (n=16). However, 11/63 cases carried variants present only in the primary or the recurrence sample and 15 carried no clearly somatic variants in either sample. There was no significant difference in the presence of matching variants between DCIS (15/30) and IBC (22/33) recurrences (p=0.28, chi-squared test). Intriguingly, we found evidence of parallel and convergent evolution, whereby the same gene was mutated in primary and recurrence but at a different site. Two clonal cases carried divergent *TP53* mutations and one had different *PIK3CA* mutations (parallel evolution), while two non-clonal cases had distinct *TBX3* and *PIK3CA* mutations respectively (convergent evolution).

Taken together, our data shows the complex genetic evolution of DCIS and its recurrences, mostly suggesting diverse evolutionary trajectories across cases and with a tendency for true invasive recurrences to be more different to their primary DCIS at a copy number level.

Clonal cases showed a significant enrichment of specific CNAs and *TP53* mutation

With the aim of predicting the likelihood of a recurrence after an initial diagnosis of DCIS, we compared recurrent cases with 32 non-recurrent cases analysed either with the targeted sequencing panel (n=18), WES (n=11) or LCWGS (n=3). Overall, the total CNA (fraction of

the genome altered) for the non-recurrent cases was lower than recurrent clonal cases (p=0.0486, Welch's t-test, Figure 4). Estimated ploidy was also higher in recurrent clonal cases compared to non-recurrent cases (p=0.024). CNA such as gain of *CCND1* (on 11q) and *ZNF217* (on 20q) genes, as well as copy number changes on 17q and 5q were more evident in clonal cases compared with non-recurrence cases (Figure 4, Supp Figure 11). However, the change in 5q may be influenced by the greater presence of ER- cases in clonal recurrences as this CNA was significantly different between ER positive clonal and ER negative clonal cases (12/44 ER+ *versus* 12/15 ER-; p=0.0006, Fisher's exact test). *PIK3CA* mutations were slightly more common in non-recurrent cases compared with clonal recurrent cases but this was not statistically significantly different (p = 0.064, Fisher's exact test, Figure 5). However, *TP53* mutations were significantly enriched in clonal compared to non-recurrent cases (p=0.009, Fisher's exact test, Figure 5).

Non-clonal cases had similar CNA and mutational status as non-recurrent except for a few changes on chromosome 8 (excluding the oncogene *MYC)* (Supp Figure 11). The only difference in clinical or histological features that indicated the likelihood of a non-clonal recurrence compared with no recurrence was that non-clonal recurrences were less likely to have had treatment with radiotherapy (p=0.014). Power was limited due to the small number of non-clonal recurrence cases: non-clonal recurrences had a lower FGA and smaller size but neither was statistically significant.

Association of TP53 with recurrence

To validate the prognostic value of *TP53* mutations, we performed TP53 IHC (Figure 5). We first evaluated the accuracy of assessing *TP53* mutation status by IHC in DCIS. There were 35 cases available with both mutation and protein data, including 11 samples with a mutation

and 24 with normal *TP53* DNA sequence. Protein staining was concordant with the mutation status in 31/35 (89%) of cases (Supplementary Figure 12). Of the three mutated cases called as normal, two could be explained by sample heterogeneity as both had mutation allele frequencies of <0.2%. The remaining case had a truncating mutation with an allele frequency of 0.71 (p.Lys132Ter) that would have been expected to lead to protein absence. One case lacking a mutation was called as complete absence staining, which potentially could be explained by poor staining performance, as very few non-tumour cells stained, and only faintly. Examples of the various staining patterns are shown in Figure 5. Overall, the concordance was sufficiently strong to proceed to a larger-scale analysis using DCIS in a tissue microarray. Cases analysed were treated with breast conserving surgery both with and without radiotherapy (RT). Abnormal TP53 staining was seen in 36/186 scorable cases (19.4%) and was associated with higher nuclear grade, negative ER and PR staining and positive HER2 staining, but not with tumour size or patient age (Supplementary Figure 12, Supplementary Table 5).

Abnormal TP53 status was significantly associated with ipsilateral recurrence in a univariate Cox regression analysis (p=0.041, HR = 1.91, 95% CI 1.03-3.55, Supplementary Figure 13). As expected, younger age was also associated with an increased risk of recurrence, while treatment with radiotherapy was associated with decreased recurrence (HR = 0.27; 95% CI 0.11-0.64). On multivariate Cox regression analysis (n=134 cases with complete data) including age group and radiotherapy in the model, TP53 status (HR = 3.13; 95% CI 1.26-7.8) and radiotherapy (HR = 0.37; 95% CI 0.14-0.99) were significantly associated with recurrence, Supplementary Figure 13). We also tested a model that included an interaction between TP53 and radiotherapy, given the known association of TP53 with the DNA damage response and radiation (44) (Figure 5). This model had a better concordance and validity

compared to the model without the interaction (C-index 0.72 vs 0.67). This interaction suggests that DCIS with abnormal TP53 does not benefit from radiotherapy in the same way that DCIS with normal TP53 does.

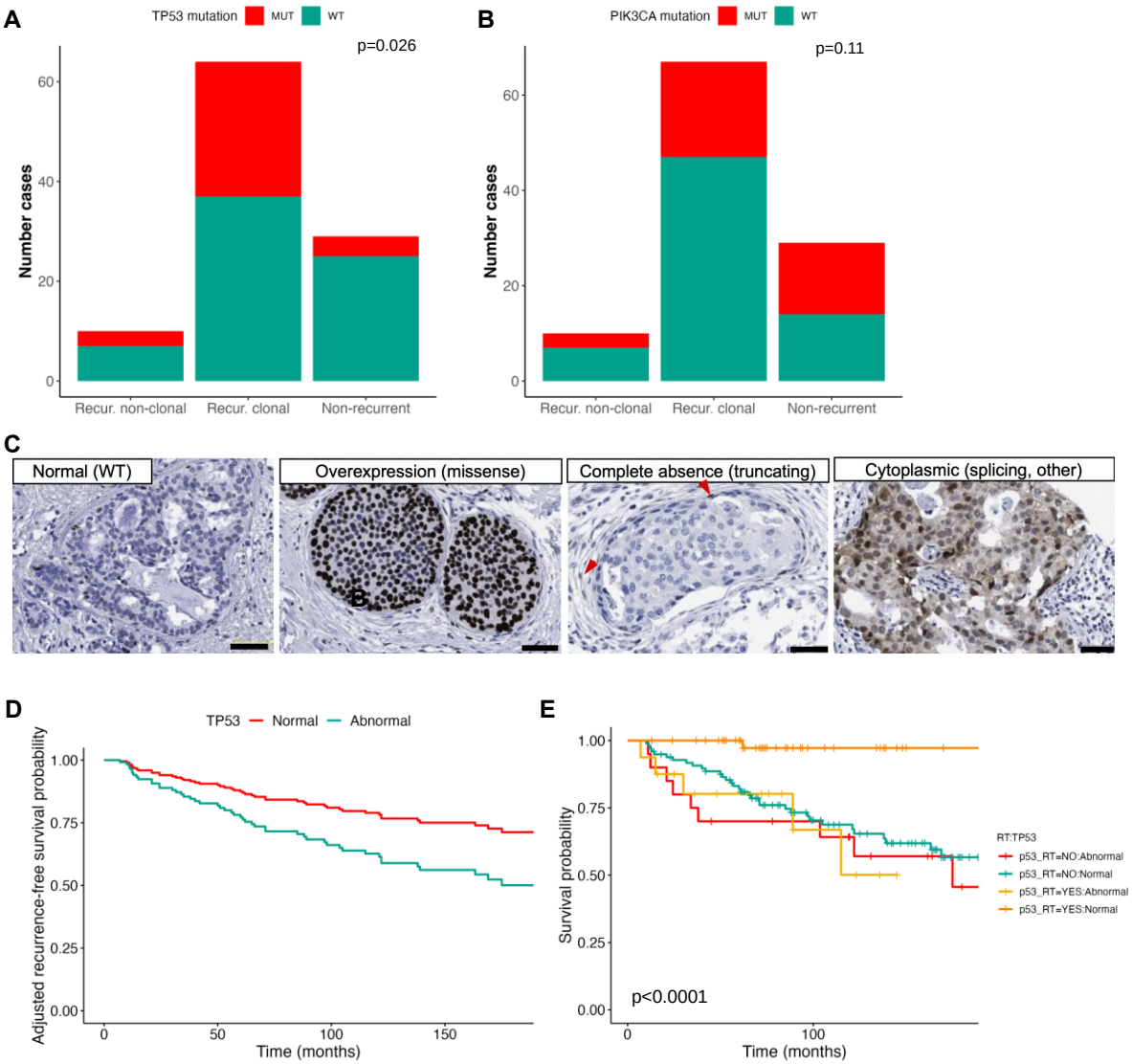


Figure 5. Association of mutations with recurrence Number of cases with **A.** *TP53* or **B.** *PIK3CA* mutation. P-value is a 2x3 Fisher's exact test comparing primary DCIS cases with a clonal recurrence, non-clonal recurrence and no recurrence. WT=Wild type, MUT=mutation. **C**. TP53 immunohistochemistry staining patterns in DCIS: normal expression with heterogeneous staining intensities in <50% of cells; overexpression pattern; complete absence of TP53 with non-tumour cells showing normal staining as an internal control (arrowheads); cytosolic staining. Black scale bar is 50 µm. **D.** Ipsilateral recurrence-free survival curves of DCIS by TP53 status, adjusted for age and radiotherapy treatment **E**. Ipsilateral recurrence-free survival curves of DCIS, illustrating the interaction between TP53

(p53) and radiotherapy (RT). The *p*-value is a likelihood ratio test. Compared to the TP53 abnormal/ radiotherapy untreated group, the TP53 normal/radiotherapy treated group had a hazard ratio of 0.045 (95% CI 0.0057- 0.356). The other comparisons were not statistically significant.

Discussion

This study challenges the previous assumption that later ipsilateral breast carcinomas after a prior DCIS are directly related (i.e. true recurrence with clonal somatic genetic events), particularly for recurrences that occur rapidly and/or are phenotypically the same (grade, hormone receptor status). Clonality implies that the recurrence has arisen from tumour cells that survived treatment. Using haplotype-specific copy number profiles we show that *de novo* breast carcinomas can occur in 14% of patients and that clonality status cannot be accurately predicted by shared grade or receptor status.

Lips and colleagues recently showed that 18% of IBC recurrences were unrelated to the primary DCIS with a large paired primary-recurrence cohort (45). Paired samples were sequenced by LCWGS and the analysis was limited by only using the total CNA profile instead of allele-specific CNA and by focussing on chromosomal changes as single events. Thus, the evolutionary history of a paired tumour was obscured and they could not detect parallel evolutionary events or WGD events. Regardless, they did confirm four cases as non-clonal using single cell DNA sequencing, reflecting independent lineages. In contrast, we generated haplotype-specific CNA by multi-sample phasing and generated MEDICC2 profiles for more than 50 DCIS-recurrence pairs. In these profiles, shared copy number segments were tracked as ancestral events and further gains or losses were detected in primary or recurrent tumours. This more powerful analysis was, for example, able to detect clonality in a case with shared ancestral CNA that was called as non-clonal by the manual breakpoint inspection method.

Clonality analyses are always limited by technological and biological resolution. We did see a decrease in the non-clonal proportion of cases when moving from lower (32%) to higher (9%) resolution sequencing. Even higher resolution analysis by WGS changed the non-clonal status of one of the four cases analysed. Notably, the majority of cases could be resolved without point mutation data. Only two cases were determined on the basis of mutation information, one targeted sequencing panel and one WES. It remains possible that the apparently independent recurrences were clonally related to the primary DCIS at a level prior to any driving somatic genetic events. The tumour could have been initiated through epigenetic alterations, and the pair could share passenger somatic variants from an early clonal expansion that we could not detect because of the lack of matching germline DNA for most cases. Indeed, non-clonal primary DCIS had many fewer CN events with 6/11 cases having <5% of the genome affected, compared to 5/67 clonal primaries and 8/32 non-recurrent DCIS. Four of the six non-clonal DCIS with low CNA also had no somatic variants. Because of this limitation, the mutational landscape remained understudied in our cohort, and in-depth phylogenetic analysis such as PyClone could not be carried out. Only known breast cancer genes, such as *PIK3CA* or *TP53* were identified and incorporated into statistical clonality tools. Similarly, the combination of few variants and FFPE-derived DNA made it inadvisable to attempt to estimate mutational signatures.

We did not investigate tumour heterogeneity in any form and we extracted DNA only from one block per sample. Therefore, we would have failed to identify any independent clones present synchronously but in a different region of the primary DCIS to that which we analysed, as previously reported for polyclonal DCIS cases (46, 47). This could be resolved in the future through multi-region sampling including 3D genome sequencing; or, in an era of

spatial biology, profiling cases by interrogating the whole tumour at very high resolution, which may indicate how non-clonal populations are spatially resolved. Nevertheless, the size of non-clonal tumours was smaller than clonal cases, suggesting this possibility would be unlikely. Previously, in a study of synchronous DCIS/IBC we identified 3/21 cases that were non-clonal. All three non-clonal DCIS were in a different block to the IBC and were contained within tumours >25 mm in size (48). Another possibility would be an existing occult "second" primary tumour elsewhere in the same breast at the time of the primary DCIS diagnosis (multi-focality) that was not detected by mammogram (49). Additional concerning foci have been reported on pre-operative magnetic resonance imaging (MRI) but missed by conventional mammogram. Due to limited breast MRI in Australia and the UK, we could not rule out this possibility. In addition to this, the precise locations of recurrent tumours were unavailable for many cases in our cohort and so the distance between them and the primary DCIS could not be investigated.

Although any recurrence is clinically important, we were particularly interested in whether the paired analysis could inform us of the genetic events underpinning the DCIS-IBC transition. This was challenging due to high inter-tumour heterogeneity, and some clonal primary-recurrent tumour pairs showed strong clonal diversification reflecting ongoing chromosomal instability. We also showed the presence of parallel evolution in clonal DCIS/IBC recurrences, an event that has been shown in many tumour types including IBC (22), which might indicate a strong microenvironmental selection pressure. Each DCIS clone we analysed might not be a direct ancestor or even close relative to the recurring IBC. Despite this complexity, we did find that invasive recurrences had higher levels of CNA, including enrichment for some chromosomal loci. Three IBC recurrent pairs showed WGD as a truncal event, with an additional two cases developing WGD. Previous studies examining

evolution of DCIS to IBC have commonly analysed synchronous tumours, which are limited by the possibility that populations of IBC cells travelled through the ducts and created "DCIS-like" ducts. For example, Pareja and colleagues made an attempt to understand the evolutionary bottleneck of DCIS progression to IBC by phylogenetic analysis of WES data in a small cohort of non-low-grade pure DCIS and DCIS synchronous with IBC (50). They and others showed that DCIS co-existing with IBC had similar genetic events including CNA but also showed evidence of WGD events in DCIS and clonal diversification (50-52). However, this study lacked recurrent samples and so couldn't capture the biology of true DCIS recurrences occurring later in time. We have now shown similar truncal WGS and clonal diversity in our time-separated lesions and with the greater power of these pairs were also able to identify CN events enriched in the IBC recurrences. The increases in specific CN events in invasive recurrences reflect the overall enrichment for some of these regions in primary DCIS with recurrences, suggesting that some cases acquired these events after surgery, or had these events present in a clone in the primary DCIS that was not sampled by our analysis. However, these differences may also reflect continued evolution of aneuploidy of these chromosome regions during the acquisition of an invasive phenotype. In our previous study of DCIS synchronous with IBC, we similarly noted an increased amplitude of CN gain in IBC regions on chromosomes 8q and 11q (48).

Our data indicates the development of independent ipsilateral tumours including as IBC after a DCIS diagnosis. The rate of independence did not seem to be related to the initial treatment options, such as radiotherapy and endocrine therapy. A hypothesis we had that breast tumours recurring after radiotherapy would be more likely to be non-clonal due to the DNA-damaging effects of RT was not supported by our data, although our power was limited (RT+/clonal:

12/62, RT+/non-clonal: 1/10) due to a relatively low proportion undergoing radiotherapy treatment (53).

The finding that 14% of so-called recurrences are actually *de novo* new primary tumours supports the idea that some patients might have a high-risk microenvironment and/or genetic predisposition for new tumour development. We were unable to identify any tumour or patient characteristics that could indicate non-clonal recurrence, although we lacked family history and other indicative epidemiological information that might inform risk. Our findings may affect management of patients with *de novo* rather than recurrent tumours, as such patients may benefit then from genetic testing or preventive therapy in the same manner as someone with a strong family history, known genetic predisposition or bilateral breast cancer (54). Indeed, *BRCA* mutation carriers have a higher risk of in-breast recurrence after a diagnosis of invasive breast cancer treated by breast conserving surgery compared to non-carriers. The risk is particularly elevated for new primary tumours and in the longer term. We detected probable germline pathogenic/likely pathogenic variants in one non-clonal case (in *ATM*) and two clonal recurrence cases (in *BRCA2* and *PALB2*). Unfortunately, the low number of cases and lack of germline data precluded establishing any increase in prevalence of germline variants in non-clonal recurrence cases.

We were able to validate the prognostic, and potentially predictive association of TP53 abnormality in an independent data set using IHC. Along with the enrichment of key CNA in the clonal recurrent cohort our findings suggest these molecular biomarkers could be used to personalise treatment. Our clonality study does raise two issues in developing such a biomarker: firstly, will the discovery and validation of tumour-intrinsic biomarkers be compromised by including non-clonal cases? Our data suggest that for strong biomarkers, it

may have a limited impact: the odds ratio for *TP53* mutations dropped from 4.49 (95% CI 1.33-19.8) to 4.2 (95% CI 1.27-18.3) when non-clonal cases were included. However, a recent study that did not control for clonality of recurrence did not see enrichment in the same CN regions (55). The second and more important implication is whether the application of a tumour-intrinsic DCIS biomarker will lead to undertreatment of the patients who are at risk of a *de novo* tumour. These patients could carry a “low risk” tumour molecular profile in their primary DCIS that might lead to a less aggressive treatment course. This impact could be alleviated by the development of complementary tumour microenvironmental biomarkers (56) as well as identification of germline risk factors through mainstreaming of pathogenic variant detection in women with DCIS. Although elements of the tumour microenvironment have previously been associated with recurrence risk, such as stromal lymphocytes and collagen structures, no studies have yet evaluated whether these are associated with clonal recurrences or *de novo* primaries.

Authors' Disclosures:

The authors declare no competing financial interests.

Authors' Contributions:

Conception and design: KLG, GBM, SBF, TK; Providing access to clinical samples: GBM, EAR, SBF, ARG, IGC, LD, SJ, SL, MC; Pathological review: JMP, EH, SIJ, NR; Performing experiments: SM, SH, HS, DL, HN, OC, MP, LT, TS, AL, ReL, DJB; TIL counts: NR, JMP; Bioinformatics support, Statistical support and data analysis: MZ, RiL, TK, SL, MJ; Phylogenetic analysis: TK, RFS, TLK; Acquisition of data, analysis and interpretation of data: TK, KLG; Identification of cases and performing additional experiments: DJB, IMM, DL, SH, EH, LD; Drafted the manuscript: TK, KLG. All authors read and approved the final manuscript; overall study supervision and ethics approval: KLG.

Acknowledgements:

This study was funded by the National Breast Cancer Foundation (NBCF IIRS-18-051). TK was supported by a Cancer Council Victoria Post-doctoral Fellowship (2020) and 2020 Priority-driven Collaborative Cancer Research Scheme and co-funded by Cancer Australia and the National Breast Cancer Foundation (2002944). This study was presented at the Press Conference of American Association of Cancer Research (AACR), at the annual meeting of 2022 by TK and TK was supported by Peter MacCallum Foundation Early Career Research Development Award (2021-2022). KLG was supported by a Victorian Cancer Agency Mid-Career Fellowship, the Peter MacCallum Foundation and Union for International Cancer Control Yamagiwa Yoshida Memorial International Study Grant. We thank the Peter MacCallum Cancer Centre Bioinformatics and Molecular Genomics core facilities, which were supported by the Australian Cancer Research Foundation and the Peter MacCallum Foundation. RFS is a Professor at the Cancer Research Center Cologne Essen (CCCE) funded

by the Ministry of Culture and Science of the State of North Rhine-Westphalia. This work was partially funded by the German Ministry for Education and Research as BIFOLD - Berlin Institute for the Foundations of Learning and Data (ref. 01IS18025A and ref 01IS18037A). We thank the technical support from the Australian Genome Research Facility (AGRF). We also thank Maria Bisignano and Stefanie Koneski from Melbourne Health Pathology Service for coordinating DCIS cases and Allan Park for database assistance for these cases. We gratefully acknowledge the assistance of Audrey Mauguen, Memorial Sloan Kettering Cancer Center for valued assistance with the Clonality package. Thanks to Claire Candido, Henry Goh, and Vanson Zhou for technical assistance. We thank the Nottingham Health Science Biobank and Breast Cancer Now Tissue Bank for the provision of tissue samples.

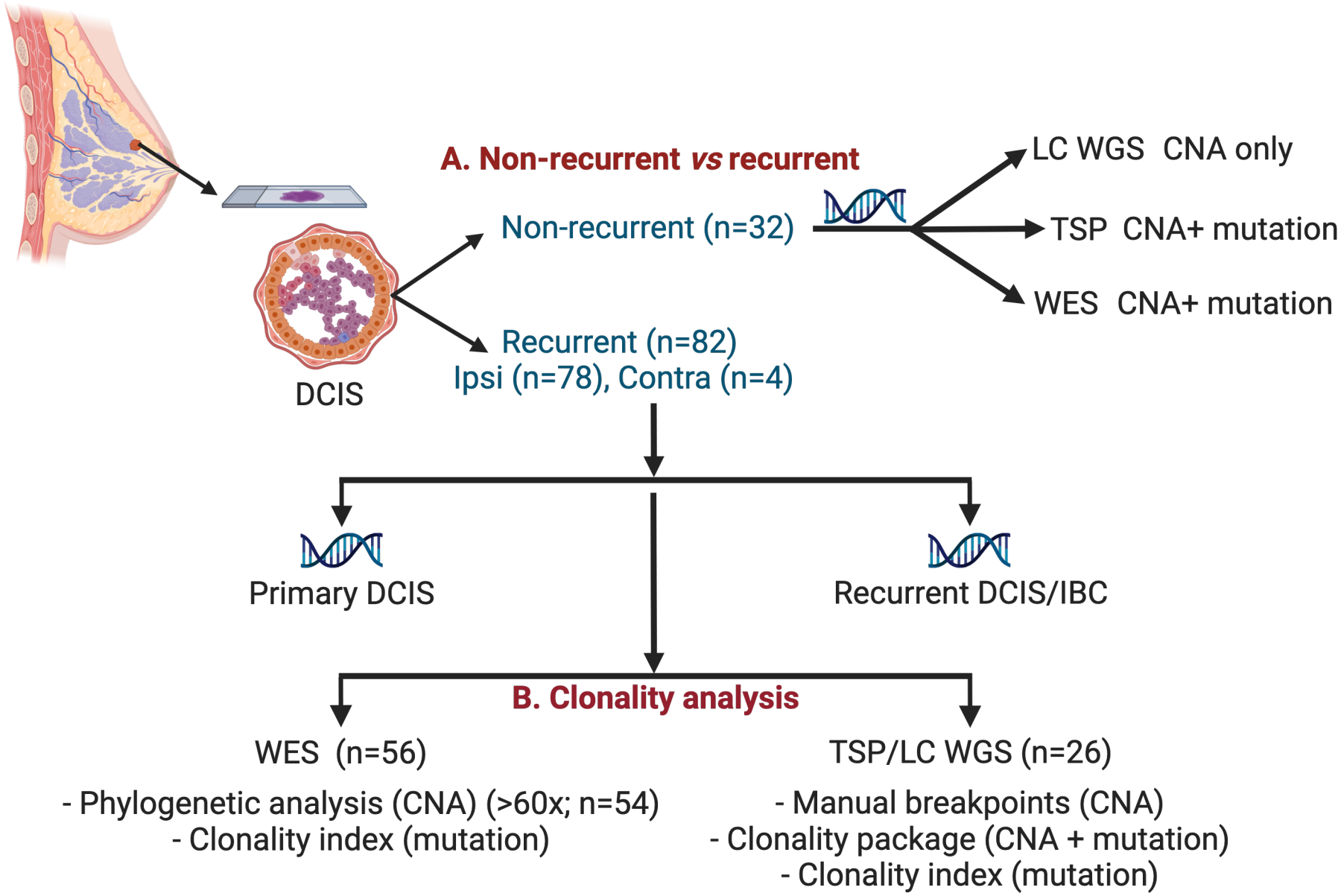
References:

1. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer Epidemiol Biomarkers Prev*. **2019**;28:835-45.
2. He W, Sofie Lindstrom L, Hall P, Czene K. Cause-specific mortality in women with breast cancer in situ. *Int J Cancer*. **2017**;140:2414-21.
3. Pinder SE, Thompson AM, Wesserling J. Low-risk DCIS. What is it? Observe or excise? *Virchows Arch*. **2022**;480:21-32.
4. Gorringe KL, Hunter SM, Pang JM, Opeskin K, Hill P, Rowley SM, et al. Copy number analysis of ductal carcinoma in situ with and without recurrence. *Mod Pathol*. **2015**;28:1174-84.
5. Lips EH, Kumar T, Megalios A, Visser LL, Sheinman M, Fortunato A, et al. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. *Nat Genet*. **2022**;54:850-60.
6. Kader T, Goode DL, Wong SQ, Connaughton J, Rowley SM, Devereux L, et al. Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue. *Genome medicine*. **2016**;8:121.
7. Kader T, Elder K, Zethoven M, Semple T, Hill P, Goode DL, et al. The genetic architecture of breast papillary lesions as a predictor of progression to carcinoma. *npj Breast Cancer*. **2020**;6:9.
8. Lee JEA, Li N, Rowley SM, Cheasley D, Zethoven M, McInerny S, et al. Molecular analysis of PALB2 associated breast cancers. *The Journal of Pathology*. **2018**.
9. Pang J-MB, Savas P, Fellowes AP, Arnau GM, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. *Modern Pathology*. **2017**.
10. Kader T, Hill P, Zethoven M, Goode DL, Elder K, Thio N, et al. Atypical ductal hyperplasia is a multipotent precursor of breast carcinoma. *The Journal of Pathology*. **2019**;248:326-38.
11. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*. **2013**.
12. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics*. **2013**:11.0. 1-.0. 33.
13. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Wilkie AO, et al. Integrating mapping-, assembly-and haplotype-based approaches for calling variants in clinical sequencing applications. *Nature genetics*. **2014**;46:912.
14. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*. **2012**;22:568-76.
15. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biology*. **2016**;17:122.
16. Pang JB, Savas P, Fellowes AP, Mir Arnau G, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. *Mod Pathol*. **2017**;30:952-63.
17. Kader T, Zethoven M, Gorringe KL. Evaluating statistical approaches to define clonal origin of tumours using bulk DNA sequencing: context is everything. *Genome Biology*. **2022**;23:1-23.
18. Schultheis AM, Ng CKY, De Filippo MR, Piscuoglio S, Macedo GS, Gatius S, et al. Massively Parallel Sequencing-Based Clonality Analysis of Synchronous Endometrioid Endometrial and Ovarian Carcinomas. *Journal of the National Cancer Institute*. **2016**;108:djv427-djv.
19. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res*. **2014**;24:2022-32.

20. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*. **2013**;14:178-92.
21. Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. *Source code for biology and medicine*. **2016**;11:1-13.
22. Watkins TBK, Lim EL, Petkovic M, Elizalde S, Birkbak NJ, Wilson GA, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature*. **2020**;587:126-32.
23. Gotovac JR, Kader T, Milne JV, Fujihara KM, Lara-Gonzalez LE, Gorringe KL, et al. Loss of SMAD4 is sufficient to promote tumorigenesis in a model of dysplastic Barrett's esophagus. *Cellular and Molecular Gastroenterology and Hepatology*. **2021**.
24. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*. **2012**;28:423-5.
25. Bollet MA, Servant N, Neuvial P, Decraene C, Lebigot I, Meyniel J-P, et al. High-resolution mapping of DNA breakpoints to define true recurrences among ipsilateral breast cancers. *JNCI: Journal of the National Cancer Institute*. **2008**;100:48-58.
26. Mauguen A, Seshan VE, Begg CB, Ostrovnaya I. Testing clonal relatedness of two tumors from the same patient based on their mutational profiles: update of the Clonality R package. *Bioinformatics*. **2019**;35:4776-8.
27. 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**;27:1698-9.
28. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. *Genome Biol*. **2022**;23:241.
29. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. *Nature*. **2013**;497:67-73.
30. Miligy IM, Gorringe KL, Toss MS, Al-Kawaz AA, Simpson P, Diez-Rodriguez M, et al. Thioredoxin-interacting protein is an independent risk stratifier for breast ductal carcinoma in situ. *Mod Pathol*. **2018**;31:1807-15.
31. Pruneri G, Lazzeroni M, Bagnardi V, Tiburzio G, Rotmensz N, DeCensi A, et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (TILs) in ductal carcinoma in situ of the breast. *Annals of Oncology*. **2016**;28:321-8.
32. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of oncology : official journal of the European Society for Medical Oncology*. **2015**;26:259-71.
33. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing Tumor-infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group Part 1 Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research. *Advances in anatomic pathology*. **2017**;24:235-51.
34. Hendry S, Pang J-MB, Byrne DJ, Lakhani SR, Cummings MC, Campbell IG, et al. Relationship of the breast ductal carcinoma in situ immune microenvironment with clinicopathological and genetic features. *Clinical Cancer Research*. **2017**.
35. Kobel M, Piskorz AM, Lee S, Lui S, LePage C, Marass F, et al. Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. *J Pathol Clin Res*. **2016**;2:247-58.
36. RCoreTeam. R: A Language and Environment for Statistical

Computing. . In: Computing RFfS, editor. Vienna, Austria: R Foundation for Statistical Computing; 2023.
37. Sjoberg D, Whiting K, Curry M, Lavery J, Larmarange J. Reproducible Summary Tables with the gtsummary Package. *The R Journal*. **2021**;13:570-80.
38. Therneau T. A Package for Survival Analysis in R. 3.5-7 ed; 2023.
39. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. **1996**;15:361-87.
40. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'. 0.4.9 ed; 2021.
41. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: Genomic Visualizations in R. *Bioinformatics*. **2016**;32:3012-4.
42. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. **2013**;6:pl1.
43. Kaufmann TL, Petkovic M, Watkins TB, Colliver EC, Laskina S, Thapa N, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. *BioRxiv*. **2021**:2021.02.28.433227.
44. Kong X, Yu D, Wang Z, Li S. Relationship between p53 status and the bioeffect of ionizing radiation. *Oncol Lett*. **2021**;22:661.
45. Lips EH, Kumar T, Megalios A, Visser LL, Sheinman M, Fortunato A, et al. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. *Nature genetics*. **2022**.
46. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell*. **2018**;172:205-17 e12.
47. Lomakin A, Svedlund J, Strell C, Gataric M, Shmatko A, Rukhovich G, et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature*. **2022**;611:594-602.
48. Johnson CE, Gorringe KL, Thompson ER, Opeskin K, Boyle SE, Wang Y, et al. Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. *Breast Cancer Res Treat*. **2012**;133:889-98.
49. Thompson JL, Wright GP. The role of breast MRI in newly diagnosed breast cancer: An evidence-based review. *Am J Surg*. **2021**;221:525-8.
50. Pareja F, Brown DN, Lee JY, Paula ADC, Selenica P, Bi R, et al. Whole-Exome Sequencing Analysis of the Progression from Non-Low Grade Ductal Carcinoma In Situ to Invasive Ductal Carcinoma. *Clinical Cancer Research*. **2020**.
51. Nachmanson D, Officer A, Mori H, Gordon J, Evans MF, Steward J, et al. The breast pre-cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ. *npj Breast Cancer*. **2022**;8:1-13.
52. Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L, et al. Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. *Nat Med*. **2017**;23:376-85.
53. Kozul C, Mann G, Silva S, Jayawardana MW, Parker B, Park A. LOCAL RECURRENCE IN DCIS TREATED PREDOMINANTLY WITHOUT RADIATION; 2022.
54. Valachis A, Nearchou AD, Lind P. Surgical management of breast cancer in BRCA-mutation carriers: a systematic review and meta-analysis. *Breast Cancer Res Treat*. **2014**;144:443-55.
55. Strand SH, Rivero-Gutierrez B, Houlahan KE, Seoane JA, King LM, Risom T, et al. Molecular classification and biomarkers of clinical outcome in breast ductal carcinoma in situ: Analysis of TBCRC 038 and RAHBT cohorts. *Cancer Cell*. **2022**;40:1521-36 e7.
56. Guerini-Rocco E, Bellerba F, Concardi A, Taormina SV, Cammarata G, Fumagalli C, et al. Expression of immune-related genes and breast cancer recurrence in women with ductal carcinoma in situ. *European Journal of Cancer*. **2024**;203:114063.

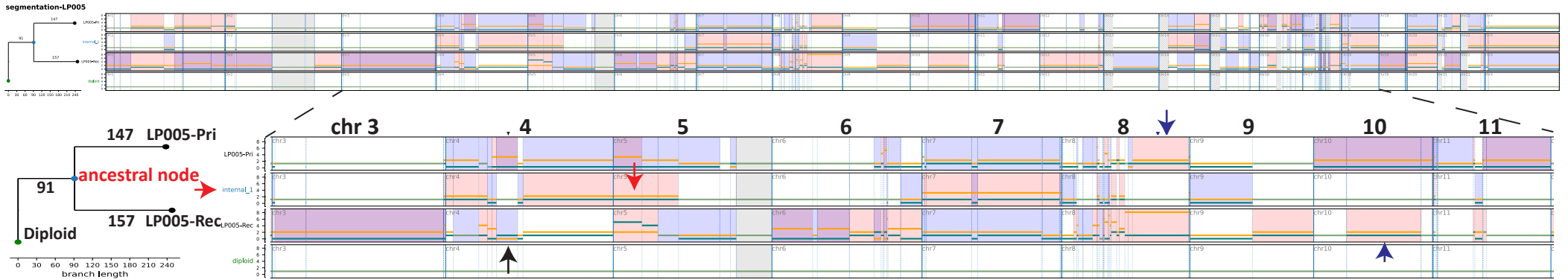
Figure 1



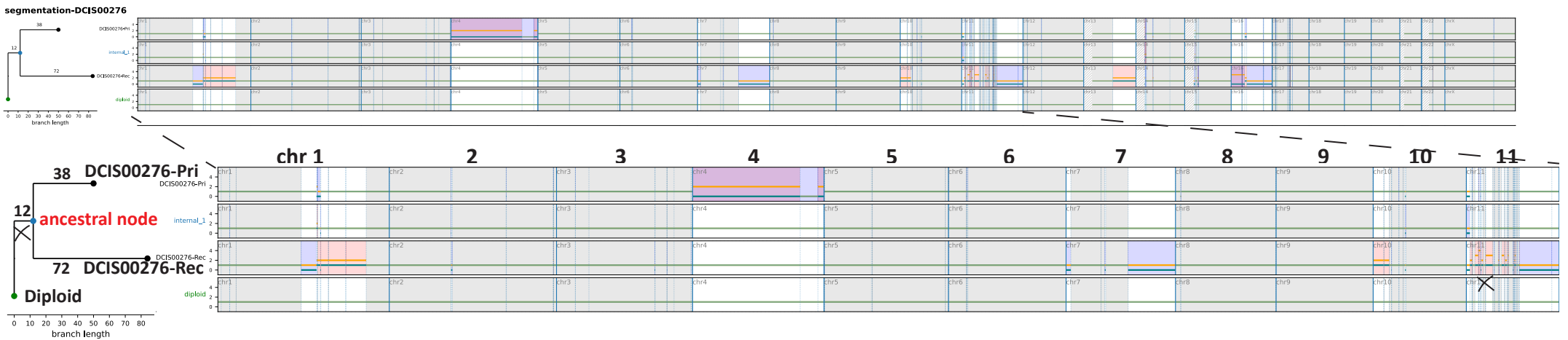
DCIS: Ductal carcinoma *in situ*; IBC: Invasive breast cancer, LC WGS: Low coverage whole genome sequencing; TSP: Targeted Sequencing Panel, WES: Whole Exome Sequencing; SCNA: Somatic copy number alterations

Figure 2

Example of a clonal pair: LP005, ER- HER2+ HG DCIS recurred as ER+ HER2+ G3 IBC



Example of a non-clonal pair: DCIS00276, ER+ HER2- DCIS recurred as ER+, HER2- IG DCIS



phased a- allele copy number

phased b- allele copy number

No copy number information

Diploid ancestor

Internal/ancestral node

Primary/Recurrent only event

Small segments, not real

Loss relative to ancestral node

Gain relative to ancestral node

Both gain & loss relative to ancestral node

Shared event

MSAI

Figure 3

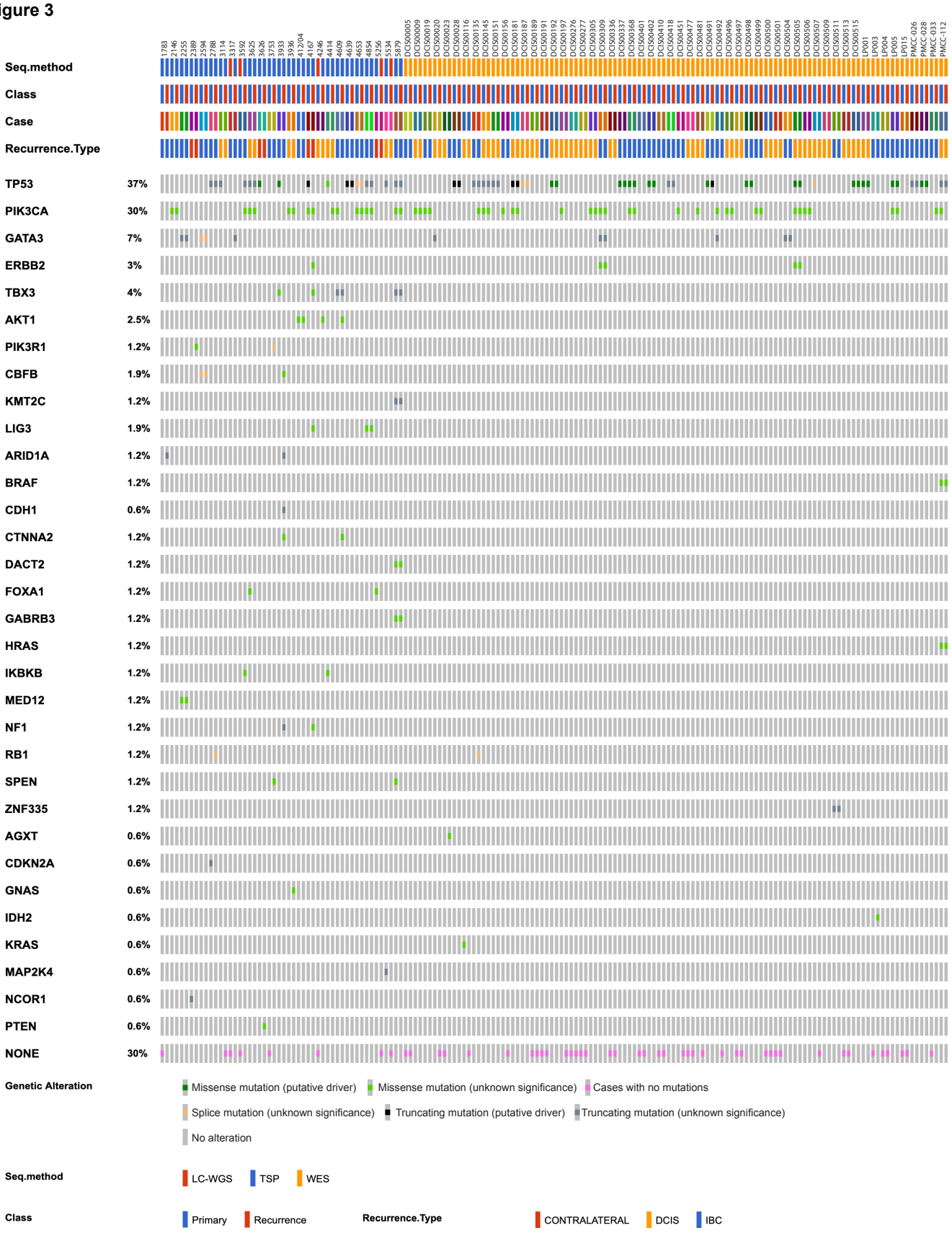

Seq.method
Class
Case
Recurrence.Type
TP53 37%
PIK3CA 30%
GATA3 7%
ERBB2 3%
TBX3 4%
AKT1 2.5%
PIK3R1 1.2%
CBFB 1.9%
KMT2C 1.2%
LIG3 1.9%
ARID1A 1.2%
BRAF 1.2%
CDH1 0.6%
CTNNA2 1.2%
DACT2 1.2%
FOXA1 1.2%
GABRB3 1.2%
HRAS 1.2%
IKBKB 1.2%
MED12 1.2%
NF1 1.2%
RB1 1.2%
SPEN 1.2%
ZNF335 1.2%
AGXT 0.6%
CDKN2A 0.6%
GNAS 0.6%
IDH2 0.6%
KRAS 0.6%
MAP2K4 0.6%
NCOR1 0.6%
PTEN 0.6%
NONE 30%
Genetic Alteration
Missense mutation (putative driver)
Missense mutation (unknown significance)
Cases with no mutations
Splice mutation (unknown significance)
Truncating mutation (putative driver)
Truncating mutation (unknown significance)
No alteration
Seq.method
LC-WGS
TSP
WES
Class
Primary
Recurrence
Recurrence.Type
CONTRALATERAL
DCIS
IBC

Figure 4

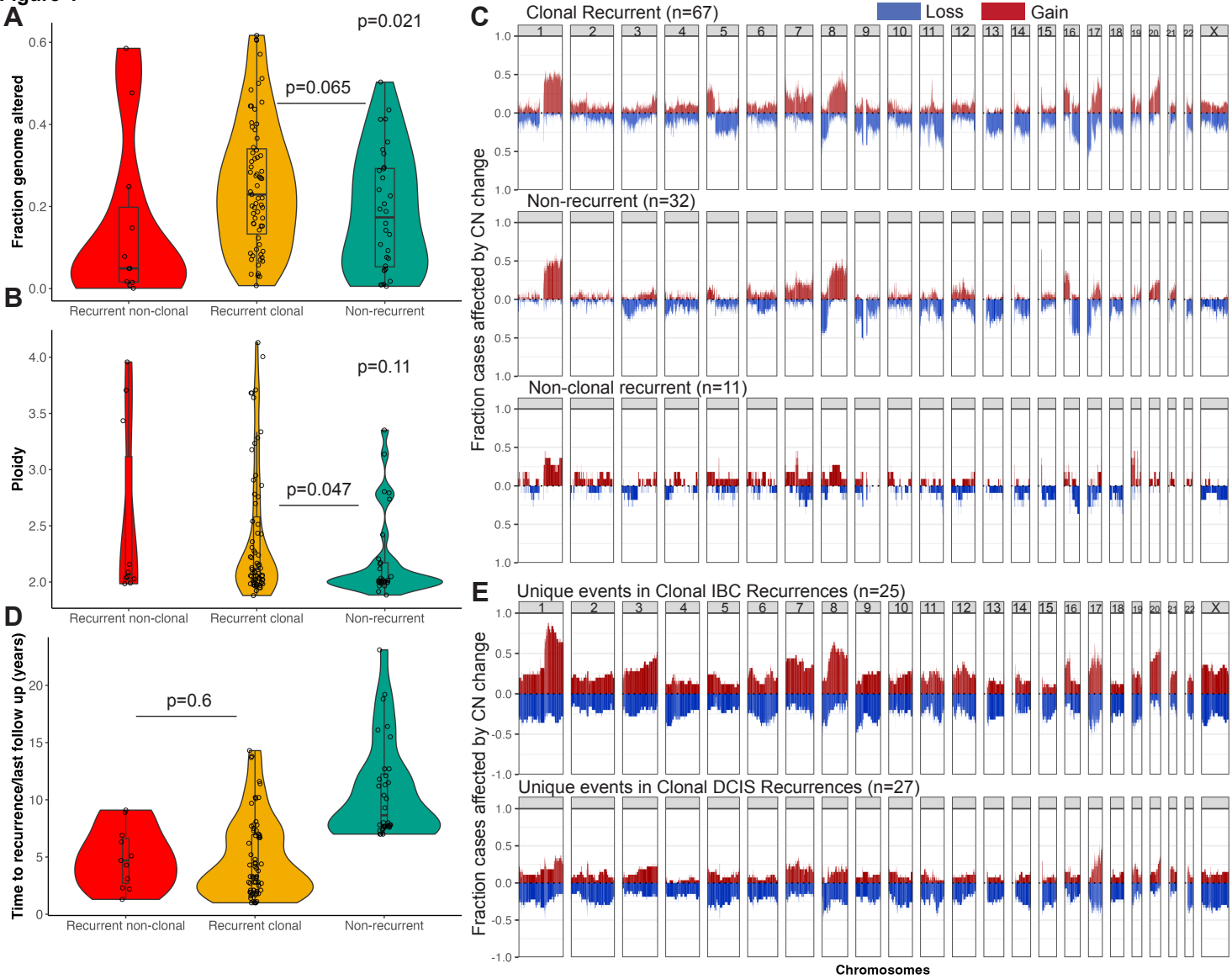
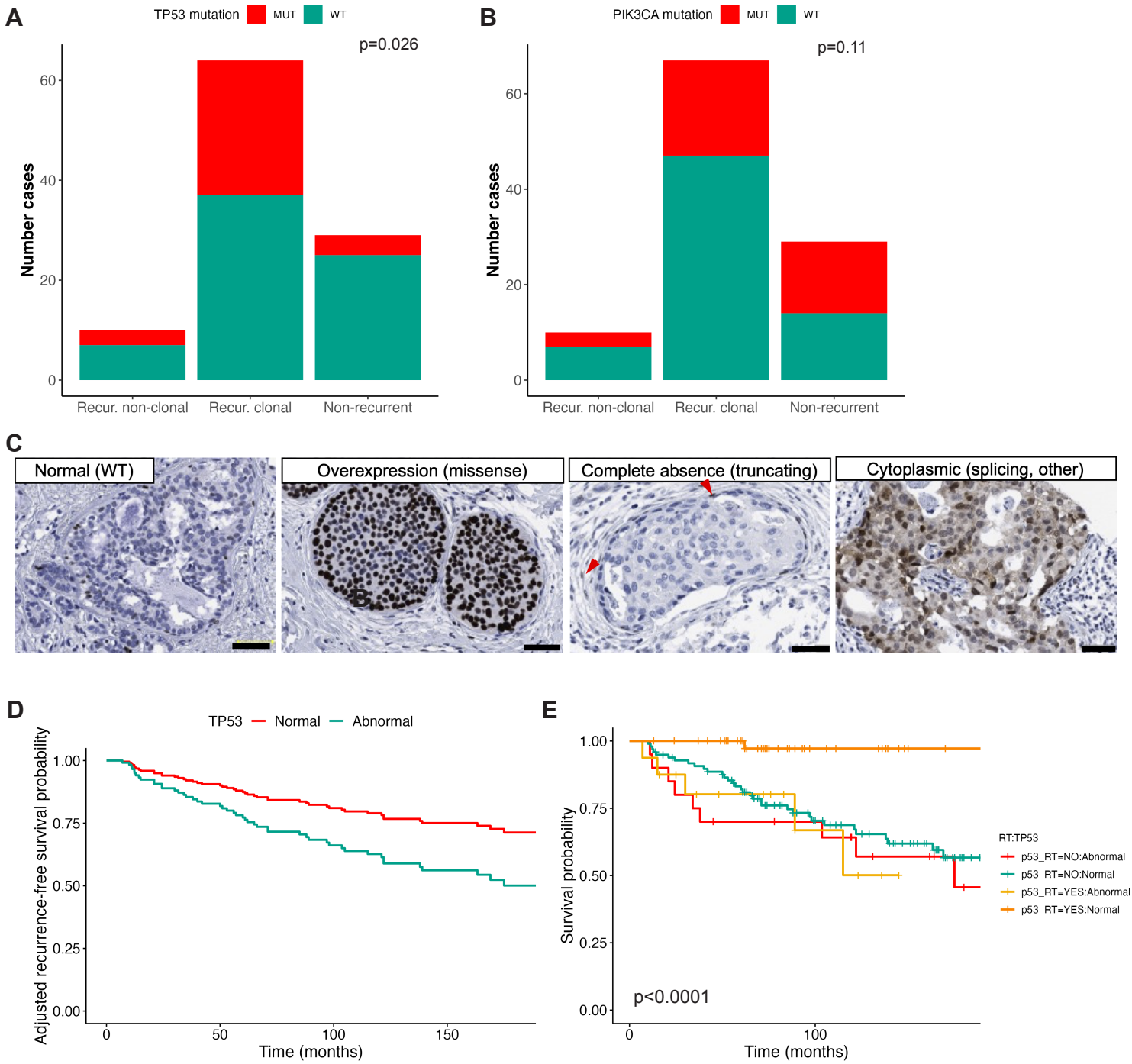

A
Fraction genome altered
p=0.021
p=0.065
Recurrent non-clonal
Recurrent clonal
Non-recurrent
B
Ploidy
p=0.11
p=0.047
Recurrent non-clonal
Recurrent clonal
Non-recurrent
D
Time to recurrence/last follow up (years)
p=0.6
Recurrent non-clonal
Recurrent clonal
Non-recurrent
C
Clonal Recurrent (n=67)
Loss
Gain
Non-recurrent (n=32)
Non-clonal recurrent (n=11)
Fraction cases affected by CN change
E
Unique events in Clonal IBC Recurrences (n=25)
Unique events in Clonal DCIS Recurrences (n=27)
Fraction cases affected by CN change
Chromosomes

Figure 5


A
TP53 mutation
MUT
WT
p=0.026
Number cases
0
20
40
60
Recur. non-clonal
Recur. clonal
Non-recurrent
B
PIK3CA mutation
MUT
WT
p=0.11
Number cases
0
20
40
60
Recur. non-clonal
Recur. clonal
Non-recurrent
C
Normal (WT)
Overexpression (missense)
Complete absence (truncating)
Cytoplasmic (splicing, other)
D
TP53
Normal
Abnormal
Adjusted recurrence-free survival probability
0.00
0.25
0.50
0.75
1.00
0
50
100
150
Time (months)
E
Survival probability
0.00
0.25
0.50
0.75
1.00
0
100
Time (months)
p<0.0001
RT:TP53
p53_RT=NO:Abnormal
p53_RT=NO:Normal
p53_RT=YES:Abnormal
p53_RT=YES:Normal

Parsed Citations

1. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. Cancer Epidemiol Biomarkers Prev. 2019;28:835-45.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

2. He W, Sofie Lindstrom L, Hall P, Czene K. Cause-specific mortality in women with breast cancer in situ. Int J Cancer. 2017;140:2414-21.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

3. Pinder SE, Thompson AM, Wesserling J. Low-risk DCIS. What is it? Observe or excise? Virchows Arch. 2022;480:21-32.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

4. Gorringe KL, Hunter SM, Pang JM, Opeskin K, Hill P, Rowley SM, et al. Copy number analysis of ductal carcinoma in situ with and without recurrence. Mod Pathol. 2015;28:1174-84.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

5. Lips EH, Kumar T, Megalios A, Visser LL, Sheinman M, Fortunato A, et al. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. Nat Genet. 2022;54:850-60.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

6. Kader T, Goode DL, Wong SQ, Connaughton J, Rowley SM, Devereux L, et al. Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue. Genome medicine. 2016;8:121.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

7. Kader T, Elder K, Zethoven M, Semple T, Hill P, Goode DL, et al. The genetic architecture of breast papillary lesions as a predictor of progression to carcinoma. npj Breast Cancer. 2020;6:9.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

8. Lee JEA, Li N, Rowley SM, Cheasley D, Zethoven M, McInerny S, et al. Molecular analysis of PALB2 associated breast cancers. The Journal of Pathology. 2018.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

9. Pang J-MB, Savas P, Fellowes AP, Arnau GM, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. Modern Pathology. 2017.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

10. Kader T, Hill P, Zethoven M, Goode DL, Elder K, Thio N, et al. Atypical ductal hyperplasia is a multipotent precursor of breast carcinoma. The Journal of Pathology. 2019;248:326-38.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

11. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997. 2013.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

12. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Current protocols in bioinformatics. 2013:11.0. 1-.0. 33.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

13. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Wilkie AO, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nature genetics. 2014;46:912.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

14. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research. 2012;22:568-76.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

15. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biology. 2016;17:122.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

16. Pang JB, Savas P, Fellowes AP, Mir Arnau G, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. Mod Pathol. 2017;30:952-63.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

17. Kader T, Zethoven M, Gorringe KL. Evaluating statistical approaches to define clonal origin of tumours using bulk DNA sequencing: context is everything. Genome Biology. 2022;23:1-23.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

18. Schultheis AM, Ng CKY, De Filippo MR, Piscuoglio S, Macedo GS, Gatius S, et al. Massively Parallel Sequencing-Based Clonality Analysis of Synchronous Endometrioid Endometrial and Ovarian Carcinomas. Journal of the National Cancer Institute. 2016;108:djv427-djv.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

19. Scheinin I, Sie D, Bengtsson H, van de Wiel MA, Olshen AB, van Thuijl HF, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. Genome Res. 2014;24:2022-32.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

20. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high- performance genomics data visualization and exploration. Briefings in bioinformatics. 2013;14:178-
Pubmed: Author and Title
Google Scholar: Google Scholar Search

92.

21. Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. Source code for biology and medicine. 2016;11:1-13.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

22. Watkins TBK, Lim EL, Petkovic M, Elizalde S, Birkbak NJ, Wilson GA, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. Nature. 2020;587:126-32.

Pubmed: Author and Title
Google Scholar: Google Scholar Search

23. Gotovac JR, Kader T, Milne JV, Fujihara KM, Lara-Gonzalez LE, Gorringe KL, et al. Loss of SMAD4 is sufficient to promote tumorigenesis in a model of dysplastic Barrett's esophagus. Cellular and Molecular Gastroenterology and Hepatology. 2021.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

24. Boeva V, Popova T, Bleakley K, Chiche P, Cappo J, Schleiermacher G, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28:423-5.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

25. Bollet MA, Servant N, Neuvial P, Decraene C, Lebigot I, Meyniel J-P, et al. High-resolution mapping of DNA breakpoints to define true recurrences among ipsilateral breast cancers. JNCI: Journal of the National Cancer Institute. 2008;100:48-58.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

26. Mauguen A, Seshan VE, Begg CB, Ostrovnaya I. Testing clonal relatedness of two tumors from the same patient based on their mutational profiles: update of the Clonality R package. Bioinformatics. 2019;35:4776-8.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

27. 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;27:1698-9.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

28. Kaufmann TL, Petkovic M, Watkins TBK, Colliver EC, Laskina S, Thapa N, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. Genome Biol. 2022;23:241.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

29. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497:67-73.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

30. Miligy IM, Gorringe KL, Toss MS, Al-Kawaz AA, Simpson P, Diez-Rodriguez M, et al. Thioredoxin-interacting protein is an independent risk stratifier for breast ductal carcinoma in situ. Mod Pathol. 2018;31:1807-15.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

31. Pruneri G, Lazzeroni M, Bagnardi V, Tiburzio G, Rotmensz N, DeCensi A, et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (TILs) in ductal carcinoma in situ of the breast. Annals of Oncology. 2016;28:321-8.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

32. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26:259-71.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

33. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing Tumor- infiltrating Lymphocytes in Solid Tumors: A Practical Review for Pathologists and Proposal for a Standardized Method From the International Immunooncology Biomarkers Working Group Part 1 Assessing the Host Immune Response, TILs in Invasive Breast Carcinoma and Ductal Carcinoma In Situ, Metastatic Tumor Deposits and Areas for Further Research. Advances in anatomic pathology. 2017;24:235-51.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

34. Hendry S, Pang J-MB, Byrne DJ, Lakhani SR, Cummings MC, Campbell IG, et al. Relationship of the breast ductal carcinoma in situ immune microenvironment with clinicopathological and genetic features. Clinical Cancer Research. 2017.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

35. Kobel M, Piskorz AM, Lee S, Lui S, LePage C, Marass F, et al. Optimized p53 immunohistochemistry is an accurate predictor of TP53 mutation in ovarian carcinoma. J Pathol Clin Res. 2016;2:247-58.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

36. RCoreTeam. R: A Language and Environment for Statistical Computing. . In: Computing RFfS, editor. Vienna, Austria: R Foundation for Statistical Computing; 2023.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

37. Sjoberg D, Whiting K, Curry M, Lavery J, Larmarange J. Reproducible Summary Tables with the gtsummary Package. The R Journal. 2021;13:570-80.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

38. Therneau T. A Package for Survival Analysis in R. 3.5-7 ed; 2023.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

39. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361-87.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

40. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using 'ggplot2'.

0.4.9 ed; 2021.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

41. Skidmore ZL, Wagner AH, Lesurf R, Campbell KM, Kunisaki J, Griffith OL, et al. GenVisR: Genomic Visualizations in R. Bioinformatics. 2016;32:3012-4.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

42. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:pl1.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

43. Kaufmann TL, Petkovic M, Watkins TB, Colliver EC, Laskina S, Thapa N, et al. MEDICC2: whole-genome doubling aware copy-number phylogenies for cancer evolution. BioRxiv. 2021:2021.02.28.433227.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

44. Kong X, Yu D, Wang Z, Li S. Relationship between p53 status and the bioeffect of ionizing radiation. Oncol Lett. 2021;22:661.

Pubmed: Author and Title
Google Scholar: Google Scholar Search

45. Lips EH, Kumar T, Megalios A, Visser LL, Sheinman M, Fortunato A, et al. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. Nature genetics. 2022.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

46. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. Cell. 2018;172:205-17 e12.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

47. Lomakin A, Svedlund J, Strell C, Gataric M, Shmatko A, Rukhovich G, et al. Spatial genomics maps the structure, nature and evolution of cancer clones. Nature. 2022;611:594-602.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

48. Johnson CE, Gorringe KL, Thompson ER, Opeskin K, Boyle SE, Wang Y, et al. Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma. Breast Cancer Res Treat. 2012;133:889-98.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

49. Thompson JL, Wright GP. The role of breast MRI in newly diagnosed breast cancer: An evidence-based review. Am J Surg. 2021;221:525-8.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

50. Pareja F, Brown DN, Lee JY, Paula ADC, Selenica P, Bi R, et al. Whole-Exome Sequencing Analysis of the Progression from Non-Low Grade Ductal Carcinoma In Situ to Invasive Ductal Carcinoma. Clinical Cancer Research. 2020.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

51. Nachmanson D, Officer A, Mori H, Gordon J, Evans MF, Steward J, et al. The breast pre- cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ. npj Breast Cancer. 2022;8:1-13.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

52. Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L, et al. Whole-genome single-cell copy number profiling from formalin-fixed paraffin-embedded samples. Nat Med. 2017;23:376-85.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

53. Kozul C, Mann G, Silva S, Jayawardana MW, Parker B, Park A. LOCAL RECURRENCE IN DCIS TREATED PREDOMINANTLY WITHOUT RADIATION; 2022.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

54. Valachis A, Nearchou AD, Lind P. Surgical management of breast cancer in BRCA-mutation carriers: a systematic review and meta-analysis. Breast Cancer Res Treat. 2014;144:443-55.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

55. Strand SH, Rivero-Gutierrez B, Houlahan KE, Seoane JA, King LM, Risom T, et al. Molecular classification and biomarkers of clinical outcome in breast ductal carcinoma in situ: Analysis of TBCRC 038 and RAHBT cohorts. Cancer Cell. 2022;40:1521-36 e7.
Pubmed: Author and Title
Google Scholar: Google Scholar Search

56. Guerini-Rocco E, Bellerba F, Concardi A, Taormina SV, Cammarata G, Fumagalli C, et al. Expression of immune-related genes and breast cancer recurrence in women with ductal carcinoma in situ. European Journal of Cancer. 2024;203:114063.

Pubmed: Author and Title
Google Scholar: Google Scholar Search