Lobular carcinomas *in situ* display intra-lesion genetic heterogeneity and clonal evolution in the progression to invasive lobular carcinoma

Ju Youn Lee[1,#], Michail Schizas[2,#], Felipe C. Geyer[1,#], Pier Selenica[1,#], Salvatore Piscuoglio[1,3], Rita A. Sakr[2], Charlotte K. Y. Ng[1,3,4], Jose V. Scarpa Carniello[2], Russell Towers[2], Dilip D. Giri[1], Victor P. de Andrade[1], Anastasios D. Papanastasiou[1], Agnes Viale[5], Reuben S. Harris[6], David B. Solit[5,7], Britta Weigelt[2,*], Jorge S. Reis-Filho[1,7*], Tari A. King[1,*]

[1]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[2]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[3]Institute of Pathology and Medical Genetics, University Hospital Basel, Basel, Switzerland

[4]Department of Biomedicine, University of Basel, Basel, Switzerland

[5]Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[6]Howard Hughes Medical Institute, Masonic Cancer Center, Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN, USA

[7]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[#]: These authors contributed equally to this work.

*: These authors jointly directed the work.

Running title: Clonal composition of LCIS and synchronous lesions

Keywords: lobular carcinoma *in situ*, invasive lobular carcinoma, whole-exome sequencing, clonal decomposition, non-obligate precursor, clonal relatedness, progression

Correspondence: Dr. Jorge S. Reis-Filho, MD PhD FRCPath, Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA. Phone: 212-639-8054. Email: reisfilj@mskcc.org; and **Dr. Tari A. King, MD**, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02215, USA. Phone: 617-632-3891 Email: tking7@bwh.harvard.edu

Conflict of interest: The authors have no conflicts of interest to declare.

Word count: 4,990; **Number of figures:** 5; **Number of tables:** 1

ABSTRACT (Word count: 248)

Purpose: Lobular carcinoma *in situ* (LCIS) is a pre-invasive lesion of the breast. We sought to define its genomic landscape, whether intra-lesion genetic heterogeneity is present in LCIS, and the clonal relatedness between LCIS and invasive breast cancers.

Experimental Design: We reanalyzed whole-exome sequencing (WES) data and performed a targeted amplicon sequencing validation of mutations identified in 43 LCIS and 27 synchronous more clinically-advanced lesions from 24 patients (nine ductal carcinomas *in situ* (DCIS), 13 invasive lobular carcinomas (ILCs) and five invasive ductal carcinomas (IDCs)). Somatic genetic alterations, mutational signatures, clonal composition and phylogenetic trees were defined using validated computational methods.

Results: WES of 43 LCIS lesions revealed a genomic profile similar to that previously reported for ILCs, with *CDH1* mutations present in 81% of the lesions. Forty-two percent (18/43) of LCIS were found to be clonally-related to synchronous DCIS and/or ILCs, with clonal evolutionary patterns indicative of clonal selection and/or parallel/branched progression. Intra-lesion genetic heterogeneity was higher among LCIS clonally-related to DCIS/ILC than in those non-clonally related to DCIS/ILC. A shift from aging to APOBEC-related mutational processes was observed in the progression from LCIS to DCIS and/or ILC in a subset of cases.

Conclusions: Our findings support the contention that LCIS has a repertoire of somatic genetic alterations similar to that of ILCs, and likely constitutes a non-obligate precursor of breast cancer. Intra-lesion genetic heterogeneity is observed in LCIS and should be considered in studies aiming to develop biomarkers of progression from LCIS to more advanced lesions.

TRANSLATIONAL RELEVANCE (Word count: 127)

We investigated the somatic genetic alterations affecting all protein coding genes in lobular carcinoma *in situ* (LCIS) and synchronously diagnosed ductal carcinomas *in situ* (DCIS) and invasive lobular (ILC) or ductal carcinomas (IDC). Our analyses revealed that LCIS is a genetically advanced lesion, often displaying intra-lesion genetic heterogeneity, with minor subclones of LCIS becoming the dominant clone in ILCs. An APOBEC-related mutational signature coupled with overexpression of *APOBEC3B* was found to be present in LCIS subclones progressing to more advanced lesions. Our findings support the notion that LCIS is a non-obligate precursor of ILC, and suggest that the development of robust molecular predictors of the risk of LCIS progression/evolution into more aggressive forms of breast cancer may benefit from the assessment of intra-lesion genetic heterogeneity in LCIS.

INTRODUCTION

Lobular carcinoma *in situ* (LCIS) is a pre-invasive lesion of the breast, which is often multifocal and bilateral (1). Over the last three decades, LCIS has been clinically perceived as a risk indicator and managed accordingly (1). There is, however, burgeoning phenotypic and genetic evidence to suggest that LCIS is a non-obligate precursor of invasive breast cancer, akin to ductal carcinoma *in situ* (DCIS) (2).

LCIS and invasive lobular carcinomas (ILCs) are phenotypically and genetically similar. Both lesions are preferentially of the luminal A molecular subtype (i.e. estrogen receptor (ER)-positive, HER2-negative, low-grade and low-proliferation), and harbor recurrent gains of 1q and losses of 16q, encompassing the *CDH1* gene locus, as well as recurrent *CDH1* somatic mutations (1,3-7). In fact, loss of E-cadherin, the protein product of the *CDH1* gene, is a hallmark feature of these lesions (3,6) and has been shown to result in the development of ILCs in conditional mouse models (8). Analyses of the genomic features of ILCs by The Cancer Genome Atlas consortium (TCGA) (6) and individual investigators (9) have revealed the genes most commonly mutated in this subtype of breast cancer, and identified molecular differences between invasive ductal carcinomas (IDCs) of no special type and ILCs, including a higher rate of *FOXA1* mutations and a lower rate of *GATA3* mutations in those with lobular histology. Additional whole-exome (WES) (7) and targeted (10) sequencing analyses focused on paired LCIS and ILCs demonstrated comparable rates of mutations affecting *CDH1*, *PIK3CA* and *CBFB*, among other genes.

Previous studies have demonstrated that synchronous LCIS and invasive breast cancers may be clonally related and share a common ancestral lesion (4,7,10). In most studies, however, clonal relatedness was inferred using limited genomic information derived from copy number (4) or targeted sequencing analyses (10). By combining copy number and WES data, Begg *et al.* provided evidence of clonal relatedness between LCIS and associated lesions (7). These studies, however, did not

investigate the basis of the clonal relatedness between LCIS and ILC, and whether the progression from LCIS to ILC would involve the selection of specific subclones or happen through multiclonal invasion (11,12). Given that not only invasive breast cancers (13) but also pre-invasive lesions (11) may be genetically heterogeneous at diagnosis, and that tumor progression/stromal invasion may stem from clonal selection (11,13), it is plausible that LCIS may display intra-lesion genetic heterogeneity and that the progression from LCIS to more clinically advanced lesions, such as DCIS or invasive breast cancer, may result from the selection of pre-existing subclones.

Here, we performed a re-analysis of WES data generated from a unique series of frozen LCIS samples from prospectively accrued, consecutive patients subjected to prophylactic or therapeutic mastectomy, previously published by Begg *et al.* (7). We performed a high-depth targeted capture sequencing validation of the mutations identified by WES in that study, using the same DNA samples and employed state-of-the-art bioinformatics algorithms with a Bayesian clustering model (PyClone) to infer subclone structure and with construction of clone based phylogeny, seeking to define the clonal composition and mutational processes in LCIS synchronously diagnosed with ILC, DCIS and/or IDC, and to ascertain whether changes in the clonal composition are observed in the progression from LCIS to DCIS or ILC.

MATERIAL AND METHODS

Subjects and samples

This study is based on the 24 cases with available WES out of the 30 cases previously subjected to microarray-based comparative genomic hybridization and/or WES by Begg *et al.* (7). Eight out of the 24 cases included in this study were also included in the targeted sequencing analysis previously reported by Sakr *et al*. (10). The cases subjected to WES include 43 LCIS and 27 synchronous more clinically-advanced lesions (**Table 1**) (Supplementary Methods).

Immunohistochemistry

Immunohistochemistry for ER, progesterone receptor (PR) and HER2 was performed essentially as previously described (11) (Supplementary Methods), and analyzed according to the American Society of Clinical Oncology (ASCO)/ College of American Pathologists (CAP) guidelines (14,15).

Whole-exome sequencing data analysis

Previously-generated (7) WES data from tumor-normal DNA samples were retrieved and re-analyzed. Neoplastic samples were sequenced to a median depth of 192x (range 95x-369x) and matched normal samples were sequenced to a median depth of 154x (range 105x-238x; **Supplementary Table S1**).

WES data analysis was performed as described in *Ng et al.* (16) and detailed in the Supplementary Methods. In brief, after aligning the reads to the reference human genome GRCh37, somatic genetic alterations were detected using state-of the-art bioinformatics algorithms and filters were subsequently applied. In addition to the identification of single nucleotide variants (SNVs) and insertions and deletions (indels), for samples from a given patient, mutations that were identified in at least one sample were subsequently interrogated in all related samples (Supplementary Methods). Given that *CDH1* germline mutations have been shown to be causative of familial gastric and breast cancer syndrome (17), the germline DNA samples from each patient were evaluated for the presence of pathogenic *CDH1* germline mutations (Supplementary Methods). The potential functional effect of each somatic mutation was defined using a combination of mutation function predictors shown to have high negative predictive value (18), as previously described (19), and genes were annotated according to their presence in three cancer gene datasets, *Kandoth et al*. (20), the Cancer Gene Census (21) and *Lawrence et al*. (22). Allele-specific copy number alterations (CNAs) and loss of heterozygosity (LOH) for specific genes were defined using FACETS (23), as previously described (16), and purity and ploidy estimations were calculated using ABSOLUTE (24) (Supplementary Methods).

Targeted amplicon re-sequencing validation of somatic mutations

A validation of the mutations found with WES was performed for cases with sufficient DNA material (n=11), using a custom designed AmpliSeq panel on an Ion Torrent Personal Genome Machine. This validation was not included in *Begg et al.* (7). Out of 4,061 somatic mutations identified by WES, 1,796 were investigated in five LCIS, five DCIS, eight ILCs, and two IDCs from cases 1, 2, 4, 7, 8, 10, 11, 12, 13, 14 and 15. 1,492 (83%) mutations were successfully validated. Mutations that had sufficient coverage in the validation experiment (minimum of 50 reads) but were not validated (allele frequency <1%) were excluded from the list of mutations used in the downstream analyses.

Clonality analysis

To infer the clonal relatedness between synchronous lesions, we defined the "clonality index" (CI) as the probability of two lesions sharing mutations not expected to have co-occurred by chance based on a previously validated method (25) (Supplementary Methods).

Clonal frequencies

To estimate the clonal architecture and composition of the lesions from each patient, mutant allelic fractions from all somatic mutations were adjusted for tumor cell content, ploidy, local copy number and sequencing errors using PyClone, as previously described (26) (Supplementary Methods).

Truncal and branch mutations

For each patient displaying at least one LCIS sample clonally-related to other lesions (LCIS, DCIS or ILC), we categorized the mutations into truncal and branch using PyClone (26) (Supplementary Methods). Truncal mutations were defined as those concurrently present in the modal populations of all LCIS and their clonally-related other lesions from a given patient. Branch mutations were defined as those comprising all non-truncal mutations.

Measure of diversity

To quantitate the intra-lesion genetic heterogeneity of each sample analyzed, we used the Shannon diversity index (27) and Gini-Simpson index (28) (Supplementary Methods).

Phylogenetic tree construction

Maximum parsimony trees were built using binary presence/absence matrices built from the somatic genetic alterations, including synonymous and non-synonymous SNVs, indels and CNAs, within the clonally-related lesions from each patient, essentially as described by *Murugaesu et al.* (16,29) (Supplementary Methods). We have also employed Treeomics as an alternative approach for the reconstruction of phylogenetic trees (30). Treeomics reconstructs phylogenies using a Bayesian inference model and determines the probability that a variant is either present or absent in a given sample.

Reverse transcription quantitative PCR (RT-qPCR)

Total RNA was extracted using TRIZOL and reverse transcribed using SuperScript VILO Master Mix (Life Technologies, Thermo Fisher Scientific) according to the manufacturer's instructions from cases for which sufficient frozen tissue samples were available. RT-qPCR was performed to analyze the expression levels of *APOBEC3B*, *APOBEC3H* and *REV1* genes using TaqMan Assay-on-Demand (Supplementary Methods).

Mutational frequencies of TCGA ILCs and luminal-A cancers

TCGA luminal-A invasive breast cancers (31) and ILCs (6) and their mutations were retrieved from the "Final Full BRCA Sample Summary" and "Mutations - Publicly accessible MAF archives" at https://tcga-data.nci.nih.gov/docs/publications/brca_2012/ and https://tcga-data.nci.nih.gov/docs/publications/brca_2015/, including all non-silent, non-RNA mutations for 209 luminal-A primary invasive breast cancers and 127 ILCs. Previous studies have demonstrated the

equivalence between the TCGA pipeline and the pipeline employed in this study for mutation detection (19,32).

Mutational signatures

To define the mutational signatures involved in the development of LCIS, DCIS and ILCs, we employed deconstructSigs (33) based on the set of mutational signatures "signature.cosmic" (34).

Statistical analysis

Analyses were performed using R. For comparisons between categorical variables, the Fisher's exact test was employed, whereas for continuous variables, the Student's t-test and Mann–Whitney U test were employed as appropriate. A hypergeometric test was performed to estimate the statistical significance of the enrichment for cancer genes (genes present in at least one of the cancer gene lists by *Lawrence et al.* (Cancer 5000-S) (22), *Kandoth et al.* (20) and/or Cancer Gene Census (21); n=745) in the genes with truncal mutations (n=559) and branch mutations (n=2,452). For the hypergeometric test, the total number of genes in the genome used was 18,986, as defined as the number of protein-coding genes by the HUGO Gene Nomenclature Committee. The representation (enrichment) factors and the *P*-values of the hypergeometric tests were provided for the analyses performed. All tests were two-sided and *P*-values<0.05 were considered statistically significant, adjusted for multiple comparisons where specified.

RESULTS

LCIS displays a repertoire of somatic genetic alterations consistent with those of ILCs and luminal A-like breast cancers

This study consists of a re-analysis of previously described WES data (7), followed by a previously unpublished targeted amplicon sequencing validation of approximately 1,800 selected mutations, from 43 LCIS and synchronous DCIS (n=9), ILCs (n=13) or IDCs (n=5) from 24 patients (**Table 1**). Three

patients underwent bilateral mastectomy, one was therapeutic for bilateral breast cancer and two patients underwent contralateral prophylactic mastectomy; these three patients were found to have bilateral LCIS (**Table 1**). All LCIS lesions were of classic type and all DCIS were of intermediate nuclear grade. For those patients with invasive lesions, tumor size and ER, PR and HER2 status in invasive tumor cells are described in **Table 1**. Notably, all invasive carcinomas were ER-positive/HER2-negative.

Somatic mutation analysis of the 43 LCIS lesions revealed a median of 20 non-synonymous somatic mutations/lesion (range 5-333) and a mutation rate of 0.39 mutations/Mb (**Figs. 1A and 2A-D)**, comparable to the number of non-synonymous somatic mutations and the mutation rates of 209 luminal-A invasive breast cancers (31) and 127 ILCs (6) from TCGA (i.e. 27 somatic mutations/lesion (range 7-203) and 0.52 mutations/Mb in luminal-A and 29 somatic mutations/lesion (range 1-1,080) and 0.56 mutations/Mb in ILCs; Mann–Whitney U test, $P>0.1$). Consistent with the notion that *CDH1* inactivation is a driver of lesions with lobular histologic features (1), we observed pathogenic mutations affecting the *CDH1* gene in 35 of 43 (81%) LCIS, of which all but three were somatic; patient 13, who had three distinct foci of LCIS, was found to harbor a *CDH1* germline mutation. All but two *CDH1* mutations were coupled with loss of heterozygosity (LOH) of the wild-type allele (77% (33/43) of all LCIS analyzed, **Fig. 1A**). Moreover, all LCIS cases lacked E-cadherin expression by immunohistochemical analysis. LCIS lacking *CDH1* mutations did not harbor mutations or deletions affecting genes coding for additional proteins that comprise the cadherin-catenin complex, such as *CTNNB1* (β-catenin), *CTNNA1* (α-catenin) or *CTNND1* (p120-catenin), nor somatic or germline genetic alterations in *RHOA* (**Supplementary Data File 1**), a gene that has been implicated in the biology of gastric cancer (35), and whose alterations result in neoplastic cells displaying discohesiveness akin to that caused by *CDH1* loss of function.

Additional genes identified by TCGA to be significantly mutated in ILCs (6), such as *PIK3CA*, *TBX3*, *FOXA1* and *MAP3K1*, were also found to be recurrently somatically mutated in LCIS (**Figs. 1A and 2E)**; however, *TP53* somatic mutations, and *PTEN* somatic mutations and homozygous deletions, present in 8%, 7%, and 6% of ILCs analyzed by TCGA (6), were not found in any of the LCIS analyzed here. Notably, *TP53* mutations were significantly more frequently found in luminal A invasive breast cancers from TCGA than in the LCIS analyzed here (12% (25/209) vs 0% (0/43), Fisher's exact test, P=0.019, **Fig. 2E**)**.** Moreover, genes identified by TCGA to be significantly mutated in luminal A invasive breast cancers, including *CBFB*, *GATA3*, *NCOR1* and *MED23* were also found be recurrently mutated in LCIS. Interestingly, however, *CBFB* was found to be mutated in 19% (8/43) of LCIS, a rate significantly higher than that in 2% (2/127) of ILCs and 2% (5/209) of luminal-A breast cancers from TCGA (Fisher's exact tests, P<0.01, **Fig. 2E**). Gene CNA analysis revealed recurrent losses of 16q and gains of 1q (**Fig. 1B**), a pattern also observed in ILCs (4,6) and luminal-A invasive breast cancers (31). Taken together, our findings demonstrate that LCIS synchronously diagnosed with more advanced lesions in this study is a genetically-advanced, neoplastic lesion often driven by E-cadherin loss of function, with a spectrum of somatic genetic alterations affecting genes commonly altered in ILCs and luminal-A invasive breast cancers.

LCIS is often clonally-related to DCIS and ILCs

WES of nine DCIS (a non-invasive precursor lesion perceived clinically to be more advanced than LCIS (36,37)), 13 ILCs, and five IDCs collected synchronously with the LCIS analyzed above demonstrated that overall these lesions displayed similar number of mutations/case, mutation rates, repertoires of CNAs and non-synonymous somatic mutations to those of the LCIS analyzed in this study (**Fig. 1, 2A-D**), with exception of *CDH1* somatic mutations that were exclusively found in LCIS and ILCs.

We reasoned that the somatic mutations and CNAs found in anatomically distinct foci of LCIS, ILC, DCIS and IDC could provide a basis for defining their clonal relatedness. Consistent with the analysis

reported by *Begg et al.* (7), but based on distinct bioinformatics and biostatistical approaches (Supplementary Methods), here we demonstrate that all multifocal LCIS originating in the same breast quadrant (8/8 samples, four patients; cases 4, 7, 9, 23) were clonally-related, harboring several identical somatic mutations and CNAs (**Fig. 3A**, **Supplementary Figs. S1-S3**). Sixty-seven percent (16/24) of multifocal LCIS affecting distinct quadrants of the breast were also clonally-related (**Fig. 3A**). Further, 10/13 (77%) ILCs and 5/9 (56%) DCIS samples were found to be clonally-related to at least one synchronous LCIS analyzed (**Fig. 3A**). Interestingly, none of the five IDCs studied were found to be clonally-related to a LCIS (**Fig. 3A**), however, in all three cases where synchronous DCIS and IDC samples were analyzed, the DCIS and IDC were found to be clonally-related (**Fig. 3A**, **Supplementary Figs. S1-S2**). As expected, no clonal relatedness was observed between lesions arising in distinct breasts (bilateral cases; **Fig. 3A, Supplementary Figs. S1-S2**). In addition, the clonal relatedness reported by *Sakr et al.* for the 8 pairs of LCIS and ILC were confirmed in this study (**Supplementary Table S2**).

Taken together, our findings indicate that the majority of multifocal LCIS lesions are clonally-related, and that the presence of these lesions in distinct quadrants of the breast does not predict their clonal relatedness. LCIS and synchronous DCIS and/or ILC are often clonally-related, corroborating the notion (4,7,38,39) that LCIS is a non-obligate precursor of more clinically-advanced lesions, in particular ILCs. Furthermore, no evidence of clonality between LCIS and IDC was observed here, suggesting that direct progression from *CDH1*-mutant LCIS to IDC is an uncommon biological phenomenon.

LCIS foci displaying intra-lesion genetic heterogeneity are more likely to progress to ILCs

Recent studies have demonstrated that intra-tumor genetic heterogeneity may be present in non-invasive lesions including DCIS (11,40) and pre-invasive lesions arising in other organs (e.g. of the esophagus)(41). In such cases, all neoplastic cells harbor the founder genetic events (i.e. truncal mutations) and subclonal populations of cancer cells display additional genetic alterations (i.e. branch

298 mutations)(42). We posited that LCIS would harbor intra-lesion genetic heterogeneity and that LCIS
299 lesions when clonally-related to DCIS or ILC would be associated with a higher level of intra-lesion
300 genetic heterogeneity than LCIS not clonally-related to more advanced lesions.
301
302 To test this hypothesis, we resolved the clonal composition of LCIS, DCIS and/or ILC samples by
303 applying a Bayesian clustering model (PyClone (26)) to mutant allele fractions, incorporating tumor
304 cellularity, ploidy and local copy number obtained from ABSOLUTE (24) and/or FACETS (23)
305 (Supplementary Methods). This analysis revealed that all but two (89%; 16/18) LCIS clonally-related to
306 DCIS/ILC but only 40% (10/25) of LCIS not clonally-related to DCIS/ILC displayed intra-lesion genetic
307 heterogeneity at the sequencing depth analyzed (Fisher's exact test, *P*=0.0016, **Supplementary Figs.**
308 **S4A-B**). These findings were further corroborated by an analysis of the Shannon and Gini-Simpson
309 diversity indices (27,28,43), which demonstrated that as a group LCIS clonally-related to DCIS and/or
310 ILC (n=18) displayed significantly higher intra-lesion genetic heterogeneity than LCIS not clonally-
311 related to more advanced lesions (n=25) (Mann–Whitney U test*, P*=0.005, **Fig. 3B-C, Supplementary**
312 **Figs. S4C-D)**. Interestingly, in case 4, composed of two LCIS and one ILC, all sharing a common
313 ancestor, the LCIS lesion displaying heterogeneity was found to be the likeliest direct precursor of the
314 ILC (**Fig. 4A**).
315
316 Given the intra-lesion genetic heterogeneity observed in LCIS, in particular in those related to more
317 advanced lesions, we sought to define whether the branch mutations found in these lesions would
318 affect 'passenger' genes or genes significantly mutated in cancer (20-22). Contrary to the notion that
319 heterogeneity would primarily affect passenger genetic events, both truncal and branch non-
320 synonymous somatic mutations detected in LCIS clonally-related to the other lesions were found to
321 target genes significantly enriched for known cancer drivers (20-22) (hypergeometric test,
322 representation factor=2.09, *P*<0.01, and hypergeometric test, representation factor=1.5, *P*<0.01,
323 respectively; **Supplementary Figs. S4E-F**). Importantly, however, in agreement with previous multi-

region analyses that suggested that most of the driver genetic alterations are early truncal events (13,16,44), the enrichment for cancer genes was higher in the constellation of truncal than in branch mutations. Truncal mutations included genes found to be significantly mutated in ILCs and/or luminal-A invasive breast cancers, including *CDH1*, *PIK3CA*, *MAP3K1*, *CBFB*, *SF3B1*, *RUNX1* and *FOXA1* (**Supplementary Data File 1**), whereas branch mutations included *GATA3*, *PIK3CA*, *ERBB2* and *KMT2C*.

Given the clonal relatedness of LCIS with DCIS and ILC, we posited that progression from LCIS to DCIS/ILC could result in the selection of specific subclones harboring private genetic alterations (11,12,40). In 29% (4/14) of cases where LCIS was clonally-related to DCIS or ILC, we observed that a selected population from the LCIS became dominant in the respective DCIS or ILC (**Fig. 4**, **Supplementary Fig. S1**), whereas in the remaining 10 cases our findings suggested parallel progression between LCIS, DCIS and/or ILC. In two cases (cases 4 and 10), a minor subclone from a LCIS was the likeliest substrate for the development of the DCIS or the ILC (**Fig. 4**). In cases 1, 11 and 16, the biological chronology of the LCIS and DCIS could not be resolved on the basis of the sequencing data available (**Supplementary Fig. S1**). Analysis of the genes affected by branch somatic mutations restricted to, or enriched in, the DCIS/ILC samples clonally-related to LCIS revealed that in the progression from LCIS to DCIS or ILC, known cancer driver genes were affected by somatic mutations (e.g. *MAP3K1* (2 cases), *RUNX1*, *NCOR1*, *ARID1A* and *TBX3* (2 cases)) or LOH of the wild-type allele (**Figs. 1 and Supplementary Fig. S4F, Supplementary Data File 1**).

Taken together, our results demonstrate that LCIS clonally-related to DCIS/ILC more frequently displays intra-lesion genetic heterogeneity than LCIS not clonally-related to more advanced lesions, that both truncal and branch mutations are enriched for known cancer drivers, and that known cancer genes are likely targeted by somatic genetic events in the progression from LCIS to more clinically advanced lesions.

Shifts in mutational processes are linked to progression from LCIS to DCIS and ILCs

There is evidence to suggest that the mutational processes that shape the mutational spectra of tumors may change during evolution (16,45). Hence, we sought to define whether changes in mutational spectra were observed in the transition from LCIS to DCIS/ILC. Given that truncal mutations are likely reflective of biological phenomena that took place prior to or during the development of LCIS, and that branch mutations in DCIS/ILC likely stem from mutational processes involved in tumor maintenance and progression, we compared the mutational spectrum of truncal and branch mutations in cases where LCIS was clonally-related to DCIS/ILC. Both truncal and branch mutations were found to be enriched for C>T transitions in the NpCpG context, consistent with a signature ascribed to aging (46), and C>G transversions and C>T transitions in the TpCpW context, suggestive of the mutational processes caused by APOBEC DNA cytosine deaminase activity (47); the latter being predominately found in the branch mutations of case 4 (**Fig. 5A**) and emerging in the DCIS of case 1 and ILC of case 18 (**Fig. 5B-C**). Akin to the variations in mutational processes observed in the progression of other cancer types (29,45), in-depth analysis of cases 1, 4 and 18 revealed that a mutational process consistent with the APOBEC signature was active in the progression from LCIS to DCIS or ILC (**Fig. 5**). Moreover, the mRNA levels of *APOBEC3B*, a DNA cytosine deaminase that has been causally implicated in the development of APOBEC signature mutations in cancer (47,48), were significantly higher in samples displaying an APOBEC mutational process than in those displaying an aging signature (**Fig. 5D**). These observations combined to indicate that, at least in a subset of cases, the APOBEC mutational process is likely to be contributing to the development of more advanced lesions.

DISCUSSION

Here we provide direct evidence of the neoplastic and non-obligate precursor nature of at least a subset of LCIS. By performing a clonal decomposition and clonal relatedness analysis of LCIS and synchronously diagnosed DCIS, ILCs and/or DCIS, we have observed that LCIS can display intra-

lesion genetic heterogeneity and be clonally-related to DCIS and ILCs, whereas progression from LCIS to IDC is likely a rare event. Notably, LCIS clonally related to ILCs and/or DCIS were found to display higher levels of intra-lesion genetic heterogeneity than LCIS that were not clonally related to a more advanced lesion, and evidence of clonal selection in the progression from LCIS to ILCs and/or DCIS was documented in a subset of patients. In these patients, the APOBEC mutational process, which has been implicated in genetic instability and intra-tumor genetic heterogeneity, appears to be present later in the evolution of LCIS and may be involved in its progression to more advanced lesions. Interestingly, the samples enriched for APOBEC mutation process displayed higher expression levels of *APOBEC3B*, whose activity has been shown to be mutagenic (47). Therefore, one hypothesize that in a subset of LCIS, upregulation of *APOBEC3B* results in increased mutagenesis and intra-tumor genetic heterogeneity, ultimately promoting subclonal expansions and progression to ILC.

LCIS has been historically considered a less advanced lesion as compared to DCIS, and is usually managed conservatively, not mandating surgical excision (1). Accordingly, in the latest version of the TNM staging system, LCIS is no longer staged as an *in situ* carcinoma (pTis) as DCIS is (49). It should be noted that although we detected clonal relatedness between LCIS and DCIS, and the LCIS as the potential substrate for the development of the DCIS (i.e. case 10), the directionality of the evolution was not clear in three cases (i.e. cases 1, 11 and 16). Hence, we cannot rule out the possibility that in a subset of cases, LCIS may have arisen from a preexistent DCIS or a common precursor (e.g. flat epithelial atypia). In fact, due to the molecular similarities between low-grade LCIS and DCIS (2), inactivation of *CDH1* in a DCIS subclone would be the likeliest explanation for such a phenotypic shift. Bi-directional progression between lesions of lobular (atypical lobular hyperplasia and LCIS) and ductal phenotype (atypical ductal hyperplasia and DCIS) is entirely consistent with the proposed concept of a low nuclear grade breast neoplasia family (2), which encompasses a group of low-grade, ER-positive neoplasms of the breast that not uncommonly affect the same segment of the breast, if not the same terminal ductal-lobular unit, and share a remarkably similar genomic landscape, having concurrent 1q

gains and 16q losses, and *PIK3CA* mutations, as their genetic signature (2). Alternatively, both LCIS and DCIS might arise from a common precursor (i.e. case 1), such as flat epithelial atypia (2). Taken together, these findings support the notion that the progression of LCIS and DCIS might be bidirectional, or that these lesions may evolve in parallel from a common ancestor.

Our findings demonstrate that LCIS displays a genomic landscape comparable to that of invasive breast cancers of luminal A subtype (31) and/or of lobular histology (6), lesions unequivocally more advanced and that mandate therapeutic intervention. Akin to ILCs (6), LCIS harbors recurrent bi-allelic inactivation of *CDH1* (77%), and recurrent mutations affecting genes commonly mutated in breast cancer, including *PIK3CA*, *FOXA1* and *TBX3*, among other genes. It should be noted, however, that genetic alterations affecting *TP53* and *PTEN*, previously found as recurrent events in ILCs (6), were not identified in the LCIS samples analyzed in this study. These differences might be related to the fact that our cohort included only classic LCIS, but given that progression may occur via clonal selection, and that not only truncal, but also branch mutations are enriched for known cancer genes, it is plausible that acquisition of genetic alterations, including those resulting in inactivation of these two *bona fide* tumor suppressor genes, may play a role in the progression to ILC. Indeed, we (7,11) and others (13) have demonstrated previously that loss of *PTEN* may be associated in the progression from DCIS to IDC.

The finding that LCIS is unlikely clonally-related to IDCs is in contrast with previous publications, including that from *Begg et al.* (7), who reported two LCIS lesions clonally-related to IDCs based on a limited number of shared mutations (one and three mutations in Patients 9 and 14, respectively), which is substantially lower than the number of shared mutations observed in clonally-related LCIS-ILC or LCIS-DCIS lesions in this study (median 12, range 2-171). The mutations described by *Begg et al.* (7) found to be shared between LCIS and IDC samples may have constituted sequencing artifacts, germline mutations or common single nucleotide polymorphisms (SNPs) (**Supplementary Table S2, Supplementary Data File 1**), as they were filtered out in our more conservative somatic mutation

analysis. Although we did not detect direct clonal relatedness between LCIS and IDC in this study, we cannot rule out the possibility that a subset of synchronous IDC and LCIS may share a common early precursor or that ductal lesions and LCIS may arise from a common earlier precursor lesion and undergo parallel evolution. In addition, it is also plausible that a subset of LCIS may stem from DCIS harboring 16q losses, but the *CDH1* inactivation takes place later in the evolution of the lesion.

Although our findings define LCIS as a non-obligate precursor of ILC, they do not imply that changes in the clinical management of patients presenting with LCIS are necessary, as the rate of subsequent breast cancer development in a large cohort of patients with a diagnosis of LCIS as reported by the SEER database demonstrates a risk of approximately 1% per year (50). Nonetheless, our study might provide a framework for the identification of markers to define LCIS cases that have a greater likelihood to progress. Although some of the pathologic characteristics of LCIS, such as volume of disease, are associated with a greater likelihood of progression to DCIS/ILC, there has yet to be a validated biomarker to predict the behavior of classic LCIS. Based on our results, one could posit that assessing the levels of intra-lesion genetic heterogeneity and/or *APOBEC3B* activity in LCIS may help select patients that should be counseled more proactively towards surgical excision and/or hormonal chemoprevention, akin to the current management of low- to intermediate-grade DCIS. Increasingly, treatment of early ER-positive breast cancer relies on pathologic features, tumor burden and genomic profiles; our findings suggest that with continued investigation a combination of clinical features, histologic classification, assessment of volume of disease and intra-lesion genetic heterogeneity may allow a more personalized risk assessment for patients with LCIS.

Our study has important limitations. The prospective accrual of frozen samples of LCIS adequate for detailed molecular studies is remarkably challenging; hence the sample size of the present study is small. In addition, our study may not be representative of incidental cases of LCIS, given that the patients included in this study were accrued in a prospective protocol for the multiregional sampling of

prophylactic and/or therapeutic mastectomies from patients with a previous diagnosis of LCIS. Moreover, we only performed WES analysis, hence we cannot rule out that non-coding alterations and/or epigenetic changes may play a role in the development and progression of LCIS. More comprehensive analyses may also be required to define the alternative drivers of *CDH1* wild-type LCIS. Finally, we used tumor bulk sequencing and state-of-the-art computational approaches to infer the clones present in each sample/case and their phylogeny. Single-cell sequencing analyses of LCIS and synchronous lesions are warranted to confirm our findings and provide direct evidence of the clonal composition of LCIS and of clonal selection in the evolution to more advanced lesions.

Despite these limitations, this proof-of-principle study demonstrates that LCIS is a neoplastic non-obligate precursor of DCIS and ILC, with a repertoire of somatic genetic alterations similar to that of ILCs and luminal-A invasive breast cancers, but lacking *TP53* and *PTEN* mutations. LCIS at diagnosis often displays intra-lesion genetic heterogeneity, and, in a subset of cases, the progression from LCIS to DCIS and ILC may involve the selection of clones, which may harbor distinct active mutational processes such as APOBEC. Our findings suggest that early documentation of intra-lesion genetic heterogeneity may be central to developing robust molecular predictors of the risk of LCIS progression/evolution into more aggressive forms of breast cancer.

DATA AVAILABILITY

WES data have been deposited in the database of Genotypes and Phenotypes (dbGaP) under the accession phs001006.v1.p1.

AUTHORS' CONTRIBUTIONS

DBS, BW, JSR-F and TAK conceived the study and supervised the work. RAS and TAK provided samples. DDG and JVSC reviewed the cases. Sample processing was performed by RAS, ADP, JVSC and RT. Massively parallel sequencing was carried out by AV. Bioinformatics analysis was performed

by JYL, MS, PS, RSL and CKYN. CKYN and JSR-F coordinated the bioinformatics analyses. JYL, MS, FCG, SP, CKYN, RAS and RSH performed data analysis and data interpretation. SP performed statistical analyses. JYL, FCG, BW and JSR-F wrote the first draft of the manuscript, which was initially reviewed by MS, SP, CKYN, RSH and TAK. All authors edited and approved the final draft.

ACKNOWLEDGMENTS: JS Reis-Filho is funded in part by the Breast Cancer Research Foundation. S Piscuoglio is funded by the Swiss National Science Foundation (Ambizione grant number PZ00P3_168165). RS Harris is the Margaret Harvey Schering Land Grant Chair for Cancer Research, a Distinguished McKnight University Professor, and an Investigator of the Howard Hughes Medical Institute. Research reported in this publication was supported in part by 2009 Komen Career Catalyst Award (T King), 2012 Komen Investigator Initiated Research Award (JS Reis-Filho), Susan G. Komen for the Cure, and by a Cancer Center Support Grant of the National Institutes of Health/National Cancer Institute (Grant No. P30CA008748). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. King TA, Reis-Filho JS. Lobular neoplasia. Surg Oncol Clin N Am **2014**;23(3):487-503 doi 10.1016/j.soc.2014.03.002.
2. Lopez-Garcia MA, Geyer FC, Lacroix-Triki M, Marchio C, Reis-Filho JS. Breast cancer precursors revisited: molecular features and progression pathways. Histopathology **2010**;57(2):171-92 doi 10.1111/j.1365-2559.2010.03568.x.
3. Weigelt B, Geyer FC, Natrajan R, Lopez-Garcia MA, Ahmad AS, Savage K*, et al.* The molecular underpinning of lobular histological growth pattern: a genome-wide transcriptomic analysis of invasive lobular carcinomas and grade- and molecular subtype-matched invasive ductal carcinomas of no special type. J Pathol **2010**;220(1):45-57 doi 10.1002/path.2629.
4. Andrade VP, Ostrovnaya I, Seshan VE, Morrogh M, Giri D, Olvera N*, et al.* Clonal relatedness between lobular carcinoma in situ and synchronous malignant lesions. Breast Cancer Res **2012**;14(4):R103 doi 10.1186/bcr3222.
5. Lu YJ, Osin P, Lakhani SR, Di Palma S, Gusterson BA, Shipley JM. Comparative genomic hybridization analysis of lobular carcinoma in situ and atypical lobular hyperplasia and potential roles for gains and losses of genetic material in breast neoplasia. Cancer Res **1998**;58(20):4721-7.
6. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A*, et al.* Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. Cell **2015**;163(2):506-19 doi 10.1016/j.cell.2015.09.033.
7. Begg CB, Ostrovnaya I, Carniello JV, Sakr RA, Giri D, Towers R*, et al.* Clonal relationships between lobular carcinoma in situ and other breast malignancies. Breast Cancer Res **2016**;18(1):66 doi 10.1186/s13058-016-0727-z.
8. Derksen PW, Liu X, Saridin F, van der Gulden H, Zevenhoven J, Evers B*, et al.* Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma

519 through induction of anoikis resistance and angiogenesis. Cancer Cell **2006**;10(5):437-49 doi
520 10.1016/j.ccr.2006.09.013.
521 9. Desmedt C, Zoppoli G, Gundem G, Pruneri G, Larsimont D, Fornili M*, et al.* Genomic
522 Characterization of Primary Invasive Lobular Breast Cancer. J Clin Oncol **2016**;34(16):1872-81
523 doi 10.1200/JCO.2015.64.0334.
524 10. Sakr RA, Schizas M, Carniello JV, Ng CK, Piscuoglio S, Giri D*, et al.* Targeted capture
525 massively parallel sequencing analysis of LCIS and invasive lobular cancer: Repertoire of
526 somatic genetic alterations and clonal relationships. Mol Oncol **2016**;10(2):360-70 doi
527 10.1016/j.molonc.2015.11.001.
528 11. Martelotto LG, Baslan T, Kendall J, Geyer FC, Burke KA, Spraggon L*, et al.* Whole-genome
529 single-cell copy number profiling from formalin-fixed paraffin-embedded samples. Nat Med
530 **2017**;23(3):376-85 doi 10.1038/nm.4279.
531 12. Casasent AK, Edgerton M, Navin NE. Genome evolution in ductal carcinoma in situ: invasion of
532 the clones. J Pathol **2017**;241(2):208-18 doi 10.1002/path.4840.
533 13. Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P*, et al.* Subclonal
534 diversification of primary breast cancer revealed by multiregion sequencing. Nat Med
535 **2015**;21(7):751-9 doi 10.1038/nm.3886.
536 14. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S*, et al.* American Society
537 of Clinical Oncology/College Of American Pathologists guideline recommendations for
538 immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin
539 Oncol **2010**;28(16):2784-95 doi 10.1200/JCO.2009.25.6529.
540 15. Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH*, et al.*
541 Recommendations for human epidermal growth factor receptor 2 testing in breast cancer:
542 American Society of Clinical Oncology/College of American Pathologists clinical practice
543 guideline update. J Clin Oncol **2013**;31(31):3997-4013 doi 10.1200/JCO.2013.50.9984.

16. Ng CKY, Bidard FC, Piscuoglio S, Geyer FC, Lim RS, de Bruijn I*, et al.* Genetic Heterogeneity in Therapy-Naive Synchronous Primary Breast Cancers and Their Metastases. Clin Cancer Res **2017**;23(15):4402-15 doi 10.1158/1078-0432.CCR-16-3115.

17. Richards FM, McKee SA, Rajpar MH, Cole TR, Evans DG, Jankowski JA*, et al.* Germline E-cadherin gene (CDH1) mutations predispose to familial gastric cancer and colorectal cancer. Hum Mol Genet **1999**;8(4):607-10.

18. Martelotto LG, Ng C, De Filippo MR, Zhang Y, Piscuoglio S, Lim R*, et al.* Benchmarking mutation effect prediction algorithms using functionally validated cancer-related missense mutations. Genome Biol **2014**;15(10):484 doi 10.1186/PREACCEPT-6413622551325626.

19. Ng CKY, Piscuoglio S, Geyer FC, Burke KA, Pareja F, Eberle CA*, et al.* The Landscape of Somatic Genetic Alterations in Metaplastic Breast Carcinomas. Clin Cancer Res **2017**;23(14):3859-70 doi 10.1158/1078-0432.CCR-16-2857.

20. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C*, et al.* Mutational landscape and significance across 12 major cancer types. Nature **2013**;502(7471):333-9 doi 10.1038/nature12634.

21. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R*, et al.* A census of human cancer genes. Nat Rev Cancer **2004**;4(3):177-83 doi 10.1038/nrc1299.

22. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR*, et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. Nature **2014**;505(7484):495-501 doi 10.1038/nature12912.

23. Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. Nucleic Acids Res **2016**;44(16):e131 doi 10.1093/nar/gkw520.

24. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T*, et al.* Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol **2012**;30(5):413-21 doi 10.1038/nbt.2203.

25. Schultheis AM, Ng CK, De Filippo MR, Piscuoglio S, Macedo GS, Gatius S*, et al.* Massively Parallel Sequencing-Based Clonality Analysis of Synchronous Endometrioid Endometrial and Ovarian Carcinomas. J Natl Cancer Inst **2016**;108(6):djv427 doi 10.1093/jnci/djv427.

26. Roth A, Khattra J, Yap D, Wan A, Laks E, Biele J*, et al.* PyClone: statistical inference of clonal population structure in cancer. Nat Methods **2014**;11(4):396-8 doi 10.1038/nmeth.2883.

27. Shannon CE. The mathematical theory of communication. 1963. MD Comput **1997**;14(4):306-17.

28. Simpson EH. Measurement of diversity. Nature **1949**.

29. Murugaesu N, Wilson GA, Birkbak NJ, Watkins T, McGranahan N, Kumar S*, et al.* Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. Cancer Discov **2015** doi 10.1158/2159-8290.CD-15-0412.

30. Reiter JG, Makohon-Moore AP, Gerold JM, Bozic I, Chatterjee K, Iacobuzio-Donahue CA*, et al.* Reconstructing metastatic seeding patterns of human cancers. Nat Commun **2017**;8:14114 doi 10.1038/ncomms14114.

31. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490(7418):61-70 doi 10.1038/nature11412.

32. Weigelt B, Bi R, Kumar R, Blecua P, Mandelker DL, Geyer FC*, et al.* The Landscape of Somatic Genetic Alterations in Breast Cancers From ATM Germline Mutation Carriers. J Natl Cancer Inst **2018** doi 10.1093/jnci/djy028.

33. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol **2016**;17:31 doi 10.1186/s13059-016-0893-4.

34. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X*, et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature **2016**;534(7605):47-54 doi 10.1038/nature17676.

595 35. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A*, et al.* Recurrent gain-of-
596 function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet **2014**;46(6):583-7 doi
597 10.1038/ng.2984.
598 36. Schnitt SJ, Allred C, Britton P, Ellis IO, Lakhani SR, Morrow M*, et al.* Ductal carcinoma in situ.
599 In: Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ, editors. WHO classification of
600 tumours of the breast. Lyon: IARC Press; 2012. p 90-4.
601 37. Vincent-Salomon A, Lucchesi C, Gruel N, Raynal V, Pierron G, Goudefroye R*, et al.* Integrated
602 genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. Clin Cancer Res
603 **2008**;14(7):1956-65 doi 10.1158/1078-0432.CCR-07-1465.
604 38. Wagner PL, Kitabayashi N, Chen YT, Shin SJ. Clonal relationship between closely
605 approximated low-grade ductal and lobular lesions in the breast: a molecular study of 10 cases.
606 Am J Clin Pathol **2009**;132(6):871-6 doi 10.1309/AJCP7AK1VWFNMCSW.
607 39. Aulmann S, Penzel R, Longerich T, Funke B, Schirmacher P, Sinn HP. Clonality of lobular
608 carcinoma in situ (LCIS) and metachronous invasive breast cancer. Breast Cancer Res Treat
609 **2008**;107(3):331-5 doi 10.1007/s10549-007-9557-0.
610 40. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W*, et al.* Multiclonal Invasion in
611 Breast Tumors Identified by Topographic Single Cell Sequencing. Cell **2018**;172(1-2):205-17
612 e12 doi 10.1016/j.cell.2017.12.007.
613 41. Maley CC, Galipeau PC, Finley JC, Wongsurawat VJ, Li X, Sanchez CA*, et al.* Genetic clonal
614 diversity predicts progression to esophageal adenocarcinoma. Nat Genet **2006**;38(4):468-73 doi
615 10.1038/ng1768.
616 42. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in
617 cancer evolution. Cancer Cell **2015**;27(1):15-26 doi 10.1016/j.ccell.2014.12.001.
618 43. Almendro V, Kim HJ, Cheng YK, Gonen M, Itzkovitz S, Argani P*, et al.* Genetic and phenotypic
619 diversity in breast tumor metastases. Cancer Res **2014**;74(5):1338-48 doi 10.1158/0008-
620 5472.CAN-13-2357-T.

44. Yates LR, Knappskog S, Wedge D, Farmery JHR, Gonzalez S, Martincorena I*, et al.* Genomic Evolution of Breast Cancer Metastasis and Relapse. Cancer Cell **2017**;32(2):169-84 e7 doi 10.1016/j.ccell.2017.07.005.

45. de Bruin EC, McGranahan N, Mitter R, Salm M, Wedge DC, Yates L*, et al.* Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science **2014**;346(6206):251-6 doi 10.1126/science.1253462.

46. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* Signatures of mutational processes in human cancer. Nature **2013**;500(7463):415-21 doi 10.1038/nature12477.

47. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC Enzymes: Mutagenic Fuel for Cancer Evolution and Heterogeneity. Cancer Discov **2015**;5(7):704-12 doi 10.1158/2159-8290.CD-15-0344.

48. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B*, et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. Nature **2013**;494(7437):366-70 doi 10.1038/nature11881.

49. Hortobagyi GN, Connolly JL, D''Orsi CJ, Edge SB, Mittendorf EA, Rugo HS*, et al.* Breast. AJCC Cancer Staging Manual. 8th ed: Springer Nature; 2017.

50. Wong SM, King T, Boileau JF, Barry WT, Golshan M. Population-Based Analysis of Breast Cancer Incidence and Survival Outcomes in Women Diagnosed with Lobular Carcinoma In Situ. Ann Surg Oncol **2017**;24(9):2509-17 doi 10.1245/s10434-017-5867-6.

TABLE 1: Clinico-pathologic characteristics of the 24 patients included in the study.

Case ID	Breast laterality	Frozen tissue blocks analyzed (n)	Type				Tumor size (invasive, mm)	Lymph node status	ER	PR	HER2	Sakr et al.	Begg et al.
			LCIS *	DCIS **	ILC	IDC							
1	Left	16	✔	✔			N/A	N/A	+	+	+	N	Y
2	Right	81	✔		✔		21	+	+	+	-	Y	Y
3	Left	52	✔	✔		✔	18	-	+	+	-	N	Y
4	Right	28	✔		✔		28	-	+	-	-	Y	Y
5	Left	27	✔		✔		23	-	+	+	-	N	Y
6	Left	28	✔		✔		16	+	+	+	-	N	Y
7	Left	27	✔	✔	✔	✔	15	-	+	+	-	N	Y
8	Left	26	✔		✔		60	+	+	+	-	Y	Y
9	Right	54	✔			✔	37	+	+	+	-	Y	Y
10	Left	24	✔	✔			N/A	N/A	+	+	-	N	Y
11	Right	97	✔	✔	✔		14	+	+	+	-	N	Y
12	Right	52	✔		✔		30	+	+	+	-	N	Y
13	Left	50	✔				N/A	N/A	+	+	-	N	Y
	Right		✔		✔		17	-	+	+	-	N	Y
14	Left	50	✔				N/A	N/A	+	+	-	N	Y
	Right		✔			✔	7.5	-	+	+	-	N	Y
15	Left	50	✔	✔		✔	18	+	+	+	-	N	Y
	Right		✔		✔		35	+	+	+	-	N	Y
16	Left	N/A	✔	✔			N/A	N/A	+	+	-	N	Y
17	Left	N/A	✔				N/A	N/A	+	+	-	N	Y
18	Left	N/A	✔		✔		16	N/A	+	+	-	Y	Y
19	Left	N/A	✔	✔			N/A	N/A	+	+	-	N	Y
20	Left	N/A	✔				N/A	N/A	+	+	-	N	Y
21	Left	N/A	✔				N/A	N/A	+	+	-	N	Y
22	Right	N/A	✔		✔		21	N/A	+	+	-	Y	Y
23	Right	N/A	✔				N/A	N/A	+	+	-	Y	Y
24	Right	N/A	✔		✔		13	N/A	+	+	-	Y	Y

DCIS, ductal carcinoma *in situ*; IDC, invasive ductal carcinoma of no special type; ILC, invasive lobular carcinoma; LCIS, lobular carcinoma *in situ*; ER, estrogen receptor; PR, progesterone receptor; HER2, epidermal growth factor receptor 2; +, positive; -, negative; N/A, not available; *all LCIS are of classic type; **all DCIS are of grade 2; Y, the case was analyzed in the previous study; N, the case was not analyzed in the previous study.

FIGURE LEGENDS

Fig. 1: Landscape of somatic genetic alterations of lobular carcinoma *in situ* (LCIS) and associated lesions.

A) Heatmap illustrating the recurrent (n≥2) somatic mutations and selected gene amplifications in LCIS (n=43), invasive lobular carcinomas (ILCs, n=13), ductal carcinoma *in situ* (DCIS, n=9) and invasive ductal carcinomas (IDC, n=5), subjected to whole-exome sequencing. Cases are shown in columns, grouped according to histologic category color-coded according to the legend, and genes in rows. Somatic mutations affecting cancer genes listed in Kandoth *et al*. (20), the Cancer Gene Census (21) and/or Lawrence *et al*. (22) are ordered from top to bottom in decreasing order of frequency, followed by selected gene amplifications. Genes highlighted in bold and/or red represent significantly mutated genes in ILCs and/or luminal-A invasive breast cancers from The Cancer Genome Atlas breast cancer studies (6,31). **B)** Heatmap illustrating the copy number alterations found in LCIS (n=43)**,** ILC (n=13), DCIS (n=9) and IDC (n=5). For A and B, mutation type, copy number states, and/or type of lesion are indicated according to the color keys on the right of the figure. Indel, insertion and deletion; SNV, single nucleotide variant.

Fig. 2: Comparison of mutation rate and frequency of mutations at gene level affecting lobular carcinoma *in situ* (LCIS) and invasive lobular carcinomas (ILCs) and Luminal A breast cancers from The Cancer Genome Atlas (TCGA) breast cancer study.

A and B) Boxplots showing the mutation burden and **C and D)** mutation rate (Mutation/Mb) in LCIS samples (n=43) and ILC (n=13), DCIS (n=9) and IDC (n=5) from this study and ILC (n=127) and Lumina A tumors (n=209) from TCGA (31). NS: not significant *p-value>0.05, **p-value>0.1 (Mann–Whitney U test). **E)** Heatmap depicting the most recurrently mutated genes affecting cancer genes identified in LCIS samples from this study and ILCs (6) and luminal A invasive breast cancers (31) from TCGA. Cases are shown in columns, genes in rows. Fisher's exact test comparisons of mutational frequencies of the mutated genes were performed between LCIS from this study (n=43) and 127 ILCs

(6) and 209 Luminal A breast cancers (31) from TCGA. The significantly different mutation frequencies between LCIS and TCGA ILCs and/or luminal A breast cancers (TCGA) are highlighted with an asterisk, where **p-value < 0.01 and *p-value < 0.05 (Fisher's exact test).

Fig. 3: Clonal relatedness and intra-lesion genetic heterogeneity in lobular carcinoma *in situ* (LCIS), invasive lobular carcinoma (ILC), ductal carcinoma *in situ* (DCIS).

A) Schematic representation of the anatomical locations (breast quadrants) of all sequenced samples in each patient and their clonal relatedness. Clonally-related lesions are connected by orange or green lines, while those in black represent lesions without a clonal relationship with any other lesion from the respective patient. In cases of unilateral LCIS, only the left or right breast was represented, and in cases of bilateral LCIS, both breasts were schematically depicted. Boxplot illustrating the distribution of **B)** the Shannon diversity index and **C)** the Gini-Simpson diversity index in LCIS not clonally-related to DCIS/ILC (n=25) and in LCIS clonally-related to DCIS and/or ILC (n=18). The colored dots indicate the cases. *P*-values of unpaired t-test with Welch's correction are indicated at the top of each figure. NS: not significant, p-value >0.05.

Fig. 4: Clonal composition of clonally-related lobular carcinoma *in situ* (LCIS) and ductal carcinoma *in situ* (DCIS) or invasive lobular carcinoma (ILC) and potential clonal selection during progression.

A and B) Decomposition of genetically distinct clones and clonal evolution in lesions from **A)** case 4 and **B)** case 10 using the results from PyClone (26). On the top left, a schematic representation of the quadrants from which each sequenced lesion was sampled is shown, and on the top right, the clonal frequency heatmap of mutations within the lesions of each case, grouped by their inferred clonal/ subclonal structure (clusters). Non-synonymous somatic mutations are shown. The clusters inferred by PyClone are shown below the clonal frequency heatmap, and the Shannon index measuring intra-lesion genetic heterogeneity for each lesion is specified within parentheses after the sample names in

the heatmap. On the bottom left, the parallel coordinates plot generated by PyClone and, in the middle right, a cluster-based phylogenetic tree based on the clusters identified by PyClone are shown. The color of the trunk and branches matches the color of their respective clusters shown in the parallel corrdinates plot. On the bottom right, a histologic lesion-based phylogenetic tree constructed using Treeomics (30) is depicted. The mutations affecting cancer genes (colored in orange) and the hotspot mutations (colored in red) that define a given clone are illustrated alongside the branches. The length of the branches is proportional to the number of mutations that distinguish a given clone from its ancestor. The numbers alongside the branches represent the total number of somatic mutations.

Fig. 5: Mutational signatures of trunk and branch mutations in lobular carcinoma *in situ* (LCIS) clonally-related ductal carcinoma *in situ* (DCIS) or invasive lobular carcinoma (ILC).

Evolution of the mutational processes with a schematic representation of the subclone structure in **A)** case 4, **B)** case 1 and **C)** case 18. Each black line represents the acquisition of somatic genetic alterations that define a given clone and each arrow depicts the divergence of a cell population from one lesion to another along with the acquisition of a set of somatic genetic alterations. The mutational signature representative of newly acquired mutations by a given subclone is depicted adjacent to each circle. The pie chart depicts the proportion of mutational signatures detected and signature 1 (aging, blue), signature 2 (APOBEC, violet) and signature 13 (APOBEC, green) are shown, with the remaining mutational signatures merged as ‘Others’ (dark gray). The number alongside the branches is the total number of somatic mutations. **D)** Reverse transcription quantitative PCR (RT-qPCR) of *APOBEC3B* (left), *APOBEC3H* (middle) and *REV1* (right) genes in samples displaying the APOBEC-related and aging-related signatures, where tissue samples were available for RNA extraction. The error bars represent the standard deviation of mean of RT-qPCR data (n=3).

Figure 1

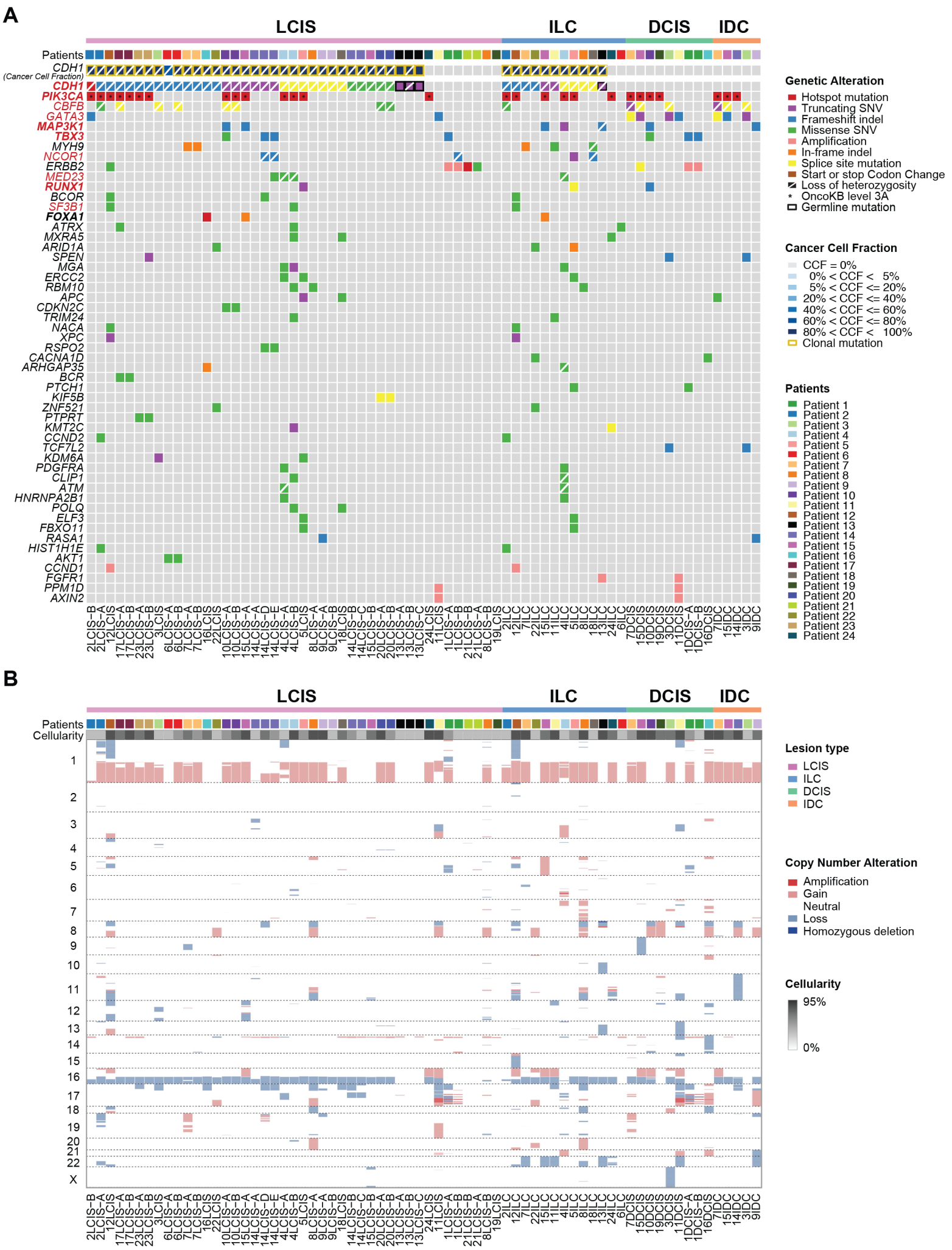
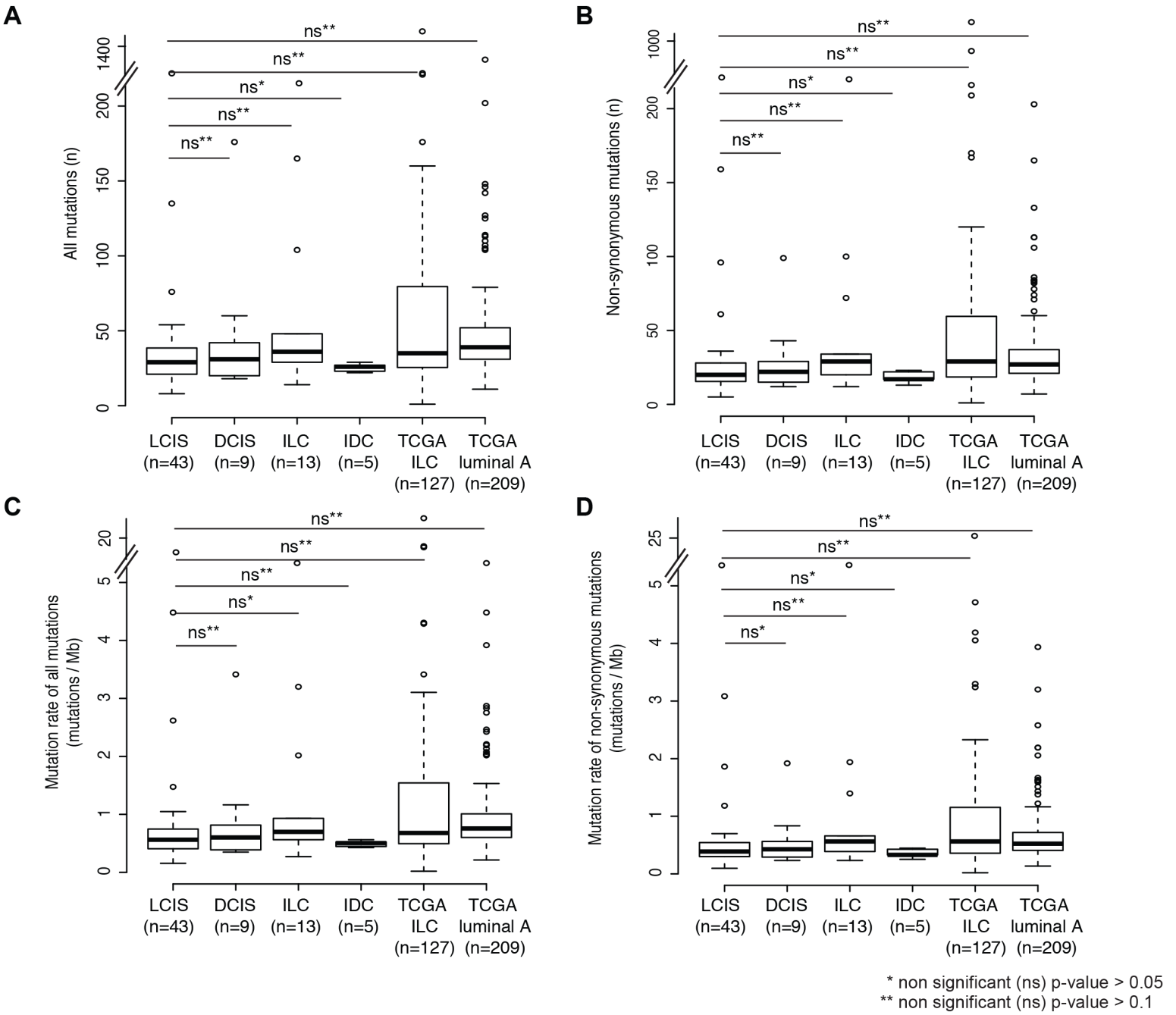

A
LCIS
ILC
DCIS
IDC
Patients
CDH1
(Cancer Cell Fraction)
CDH1
PIK3CA
CBFB
GATA3
MAP3K1
TBX3
MYH9
NCOR1
ERBB2
MED23
RUNX1
BCOR
SF3B1
FOXA1
ATRX
MXRA5
ARID1A
SPEN
MGA
ERCC2
RBM10
APC
CDKN2C
TRIM24
NACA
XPC
RSPO2
CACNA1D
ARHGAP35
BCR
PTCH1
KIF5B
ZNF521
PTPRT
KMT2C
CCND2
TCF7L2
KDM6A
PDGFRA
CLIP1
ATM
HNRNPA2B1
POLQ
ELF3
FBXO11
RASA1
HIST1H1E
AKT1
CCND1
FGFR1
PPM1D
AXIN2
Genetic Alteration
Hotspot mutation
Truncating SNV
Frameshift indel
Missense SNV
Amplification
In-frame indel
Splice site mutation
Start or stop Codon Change
Loss of heterozygosity
OncoKB level 3A
Germline mutation
Cancer Cell Fraction
CCF = 0%
0% < CCF < 5%
5% < CCF <= 20%
20% < CCF <= 40%
40% < CCF <= 60%
60% < CCF <= 80%
80% < CCF < 100%
Clonal mutation
Patients
Patient 1
Patient 2
Patient 3
Patient 4
Patient 5
Patient 6
Patient 7
Patient 8
Patient 9
Patient 10
Patient 11
Patient 12
Patient 13
Patient 14
Patient 15
Patient 16
Patient 17
Patient 18
Patient 19
Patient 20
Patient 21
Patient 22
Patient 23
Patient 24
B
LCIS
ILC
DCIS
IDC
Patients
Cellularity
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Lesion type
LCIS
ILC
DCIS
IDC
Copy Number Alteration
Amplification
Gain
Neutral
Loss
Homozygous deletion
Cellularity
95%
0%

Figure 2


A
All mutations (n)
B
Non-synonymous mutations (n)
C
Mutation rate of all mutations (mutations / Mb)
D
Mutation rate of non-synonymous mutations (mutations / Mb)
ns**
ns*
LCIS (n=43)
DCIS (n=9)
ILC (n=13)
IDC (n=5)
TCGA ILC (n=127)
TCGA luminal A (n=209)
* non significant (ns) p-value > 0.05
** non significant (ns) p-value > 0.1

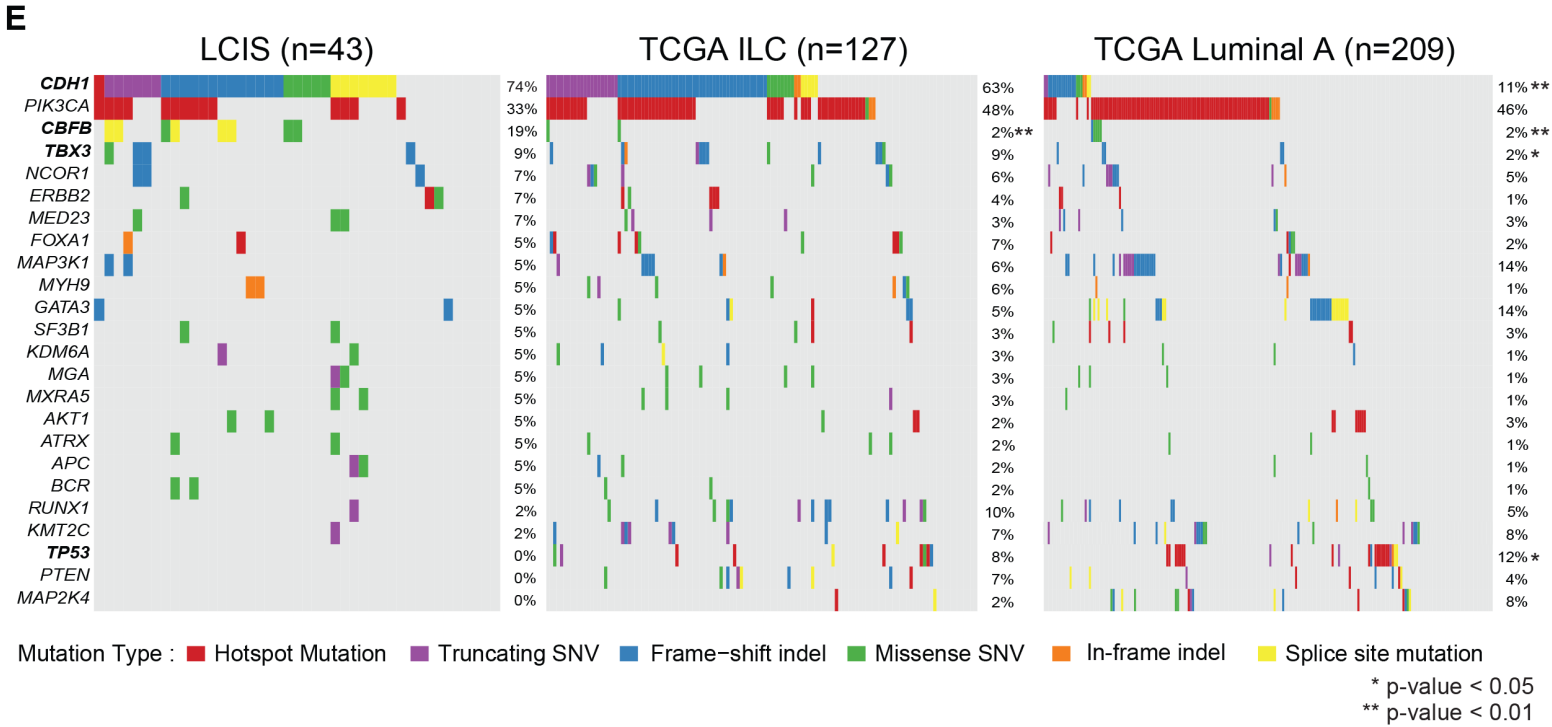
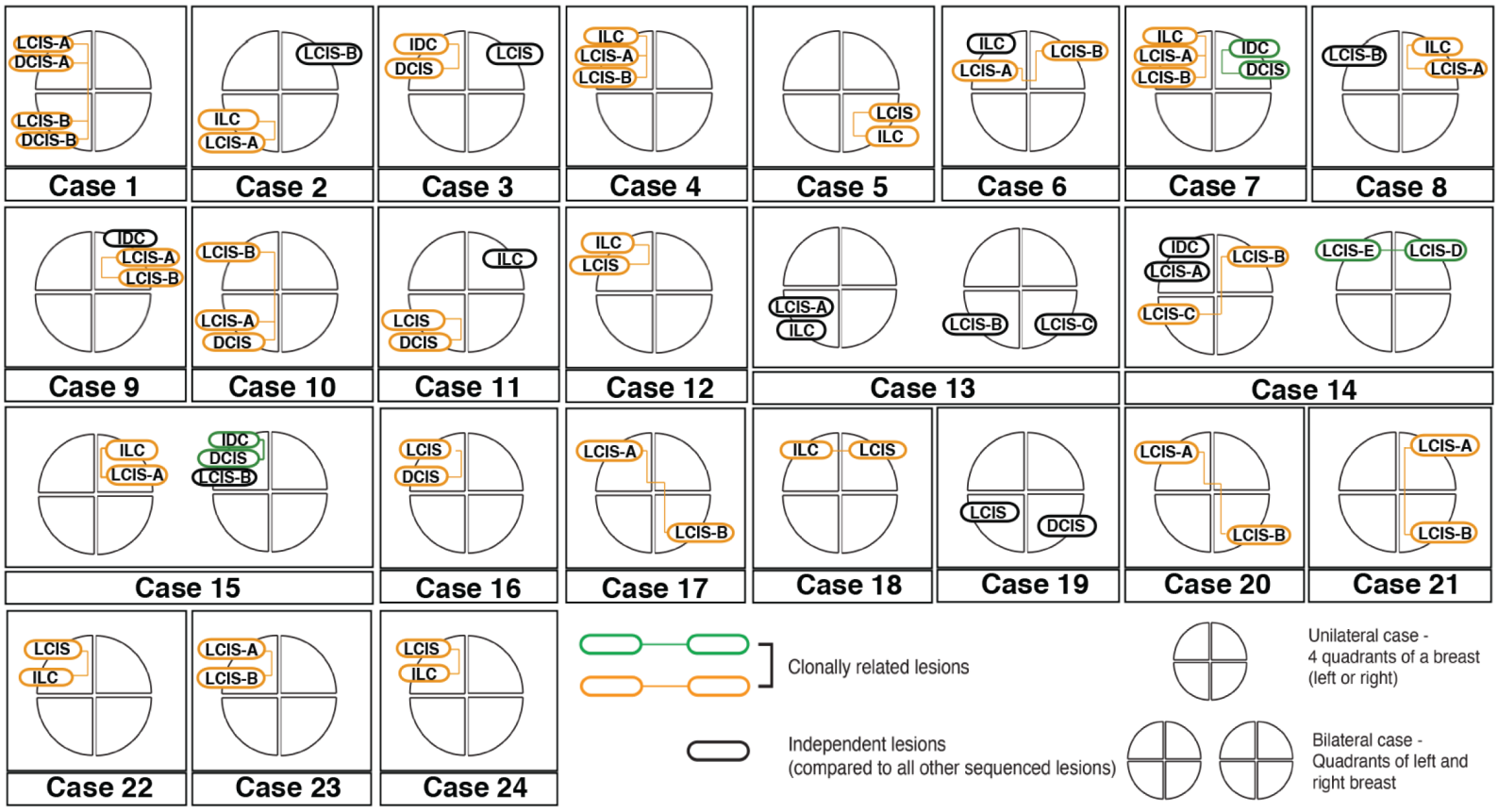
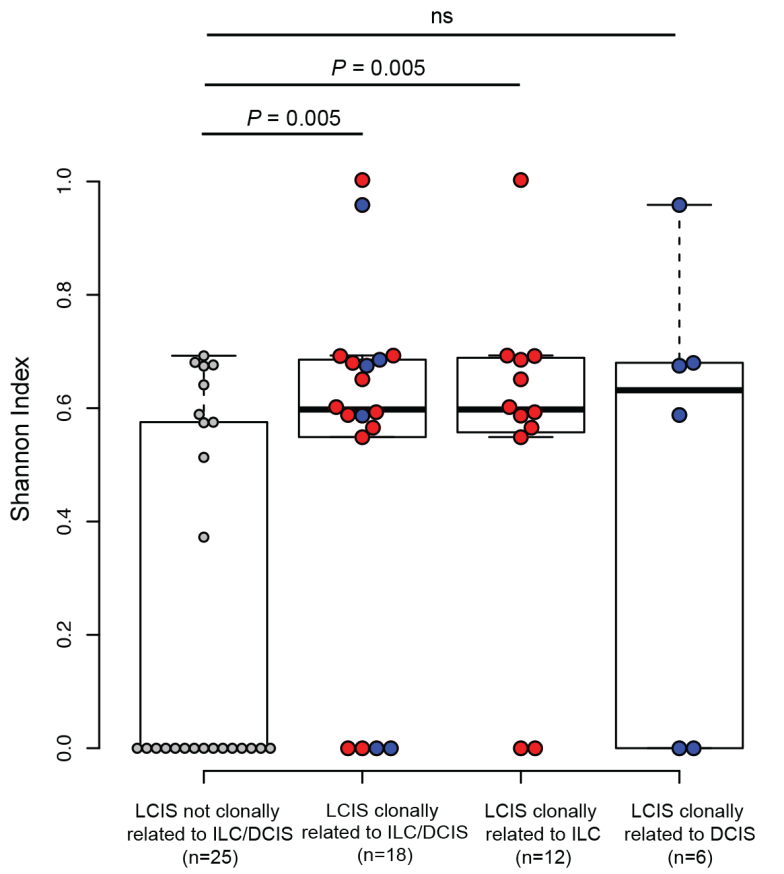

E
LCIS (n=43)
TCGA ILC (n=127)
TCGA Luminal A (n=209)
CDH1
PIK3CA
CBFB
TBX3
NCOR1
ERBB2
MED23
FOXA1
MAP3K1
MYH9
GATA3
SF3B1
KDM6A
MGA
MXRA5
AKT1
ATRX
APC
BCR
RUNX1
KMT2C
TP53
PTEN
MAP2K4
Mutation Type : Hotspot Mutation Truncating SNV Frame−shift indel Missense SNV In-frame indel Splice site mutation
* p-value < 0.05
** p-value < 0.01

Figure 3

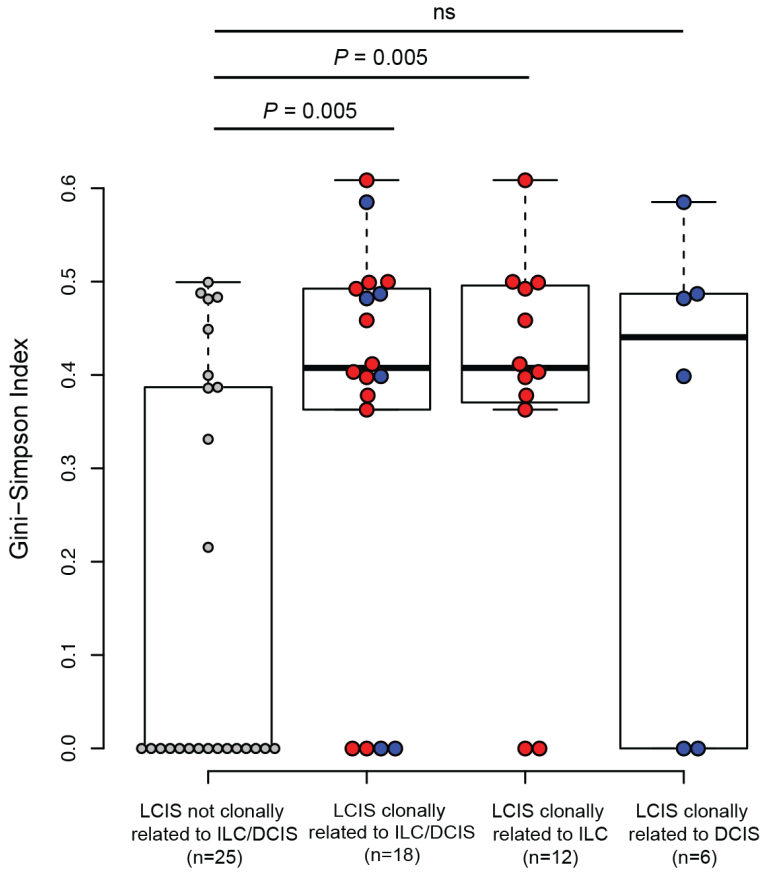
A



B



C

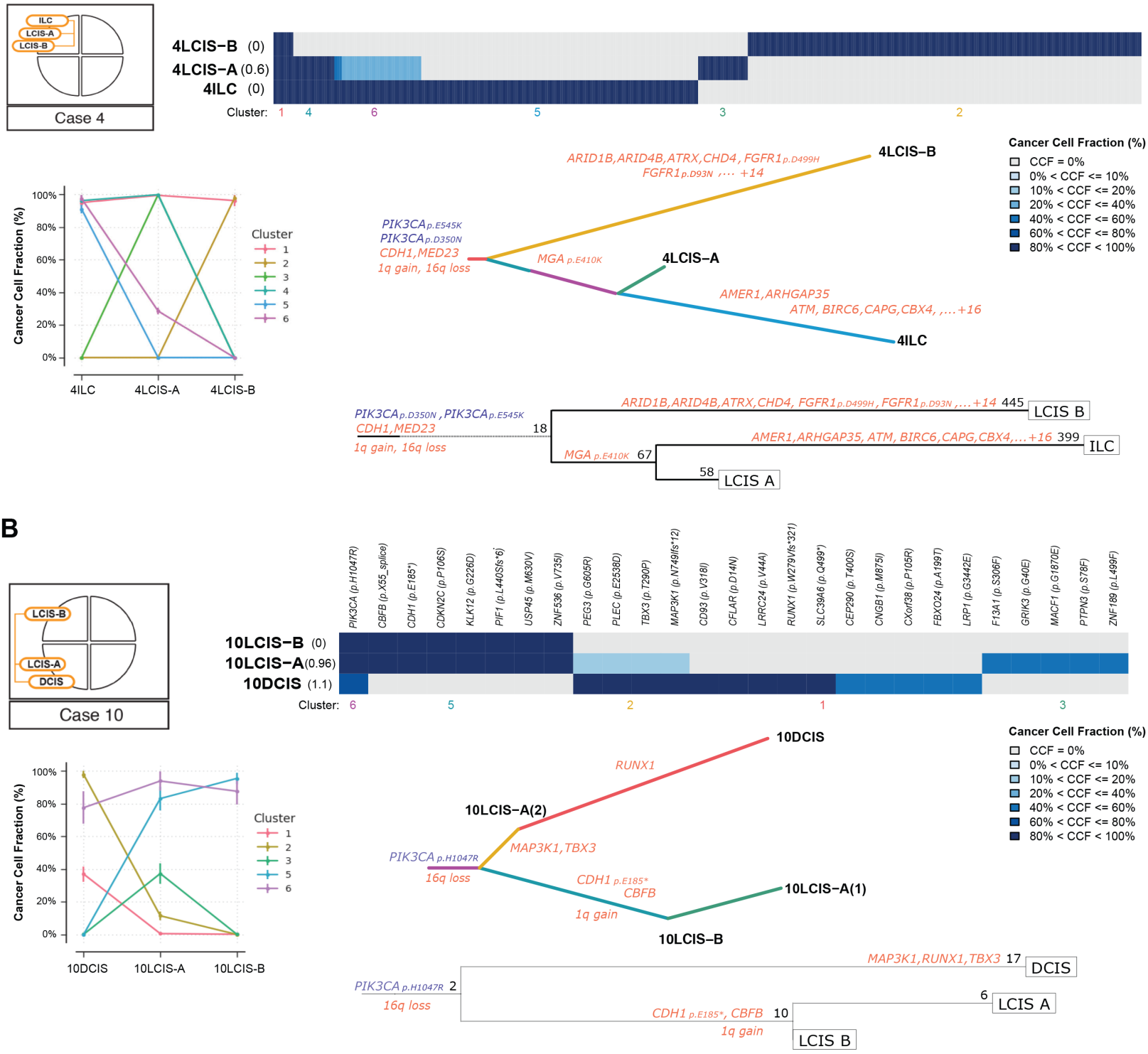


LCIS not clonally related to ILC/DCIS LCIS clonally related to ILC LCIS clonally related to DCIS

non significant (ns) p-value > 0.05

Figure 4

A

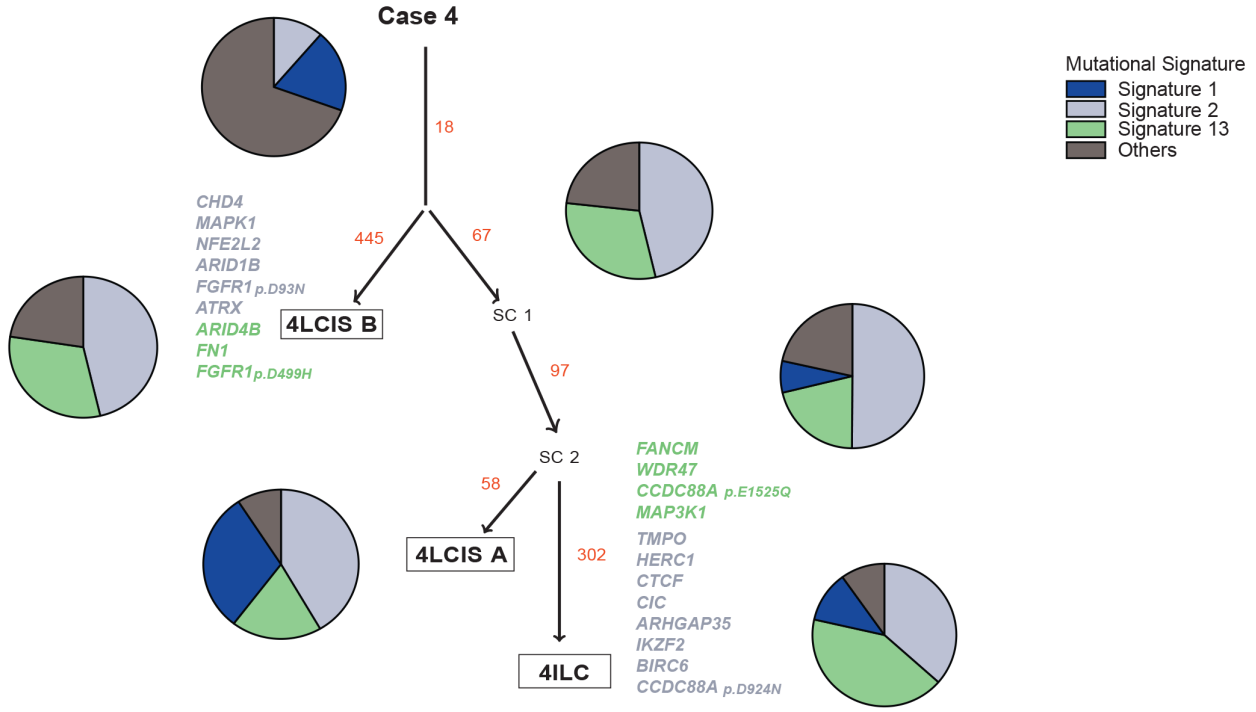

ILC
LCIS-A
LCIS-B
Case 4
4LCIS−B (0)
4LCIS−A (0.6)
4ILC (0)
Cluster: 1 4 6 5 3 2
Cancer Cell Fraction (%)
100%
80%
60%
40%
20%
0%
4ILC
4LCIS-A
4LCIS-B
Cluster
1
2
3
4
5
6
ARID1B,ARID4B,ATRX,CHD4, FGFR1p.D499H
FGFR1p.D93N ,... +14
4LCIS−B
PIK3CAp.E545K
PIK3CAp.D350N
CDH1,MED23
1q gain, 16q loss
MGA p.E410K
4LCIS−A
AMER1,ARHGAP35
ATM, BIRC6,CAPG,CBX4, ,...+16
4ILC
Cancer Cell Fraction (%)
CCF = 0%
0% < CCF <= 10%
10% < CCF <= 20%
20% < CCF <= 40%
40% < CCF <= 60%
60% < CCF <= 80%
80% < CCF < 100%
PIK3CAp.D350N , PIK3CAp.E545K
CDH1,MED23
1q gain, 16q loss
18
ARID1B,ARID4B,ATRX,CHD4, FGFR1p.D499H, FGFR1p.D93N ,...+14 445
LCIS B
AMER1,ARHGAP35, ATM, BIRC6,CAPG,CBX4,...+16 399
ILC
MGA p.E410K 67
58
LCIS A

B

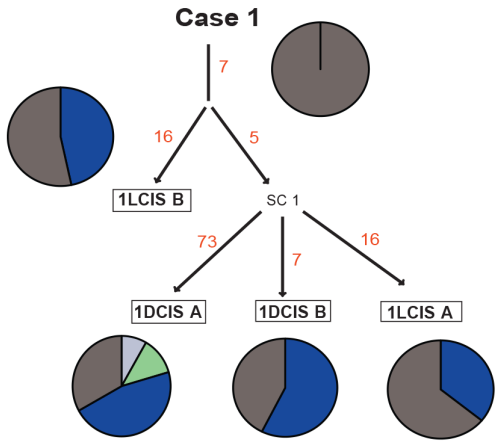
LCIS-B
LCIS-A
DCIS
Case 10
PIK3CA (p.H1047R)
CBFB (p.X55_splice)
CDH1 (p.E185*)
CDKN2C (p.P106S)
KLK12 (p.G226D)
PIF1 (p.L440Sfs*6)
USP45 (p.M630V)
ZNF536 (p.V735I)
PEG3 (p.G605R)
PLEC (p.E2538D)
TBX3 (p.T290P)
MAP3K1 (p.N749Ifs*12)
CD93 (p.V318I)
CFLAR (p.D14N)
LRRC24 (p.V44A)
RUNX1 (p.W279Vfs*321)
SLC39A6 (p.Q499*)
CEP290 (p.T400S)
CNGB1 (p.M875I)
CXorf38 (p.P105R)
FBXO24 (p.A199T)
LRP1 (p.G3442E)
F13A1 (p.S306F)
GRIK3 (p.G40E)
MACF1 (p.G1870E)
PTPN3 (p.S78F)
ZNF189 (p.L499F)
10LCIS−B (0)
10LCIS−A(0.96)
10DCIS (1.1)
Cluster: 6 5 2 1 3
Cancer Cell Fraction (%)
100%
80%
60%
40%
20%
0%
10DCIS
10LCIS-A
10LCIS-B
Cluster
1
2
3
5
6
10DCIS
RUNX1
10LCIS−A(2)
PIK3CA p.H1047R
16q loss
MAP3K1,TBX3
CDH1 p.E185*
CBFB
1q gain
10LCIS−A(1)
10LCIS−B
Cancer Cell Fraction (%)
CCF = 0%
0% < CCF <= 10%
10% < CCF <= 20%
20% < CCF <= 40%
40% < CCF <= 60%
60% < CCF <= 80%
80% < CCF < 100%
MAP3K1,RUNX1,TBX3 17
DCIS
PIK3CA p.H1047R 2
16q loss
6
LCIS A
CDH1 p.E185*, CBFB 10
1q gain
LCIS B

Figure 5

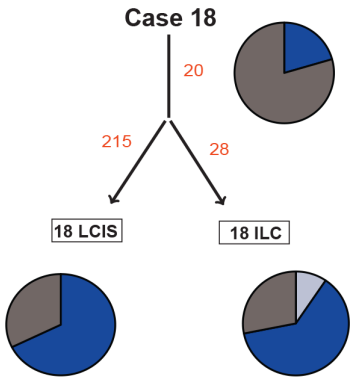
A



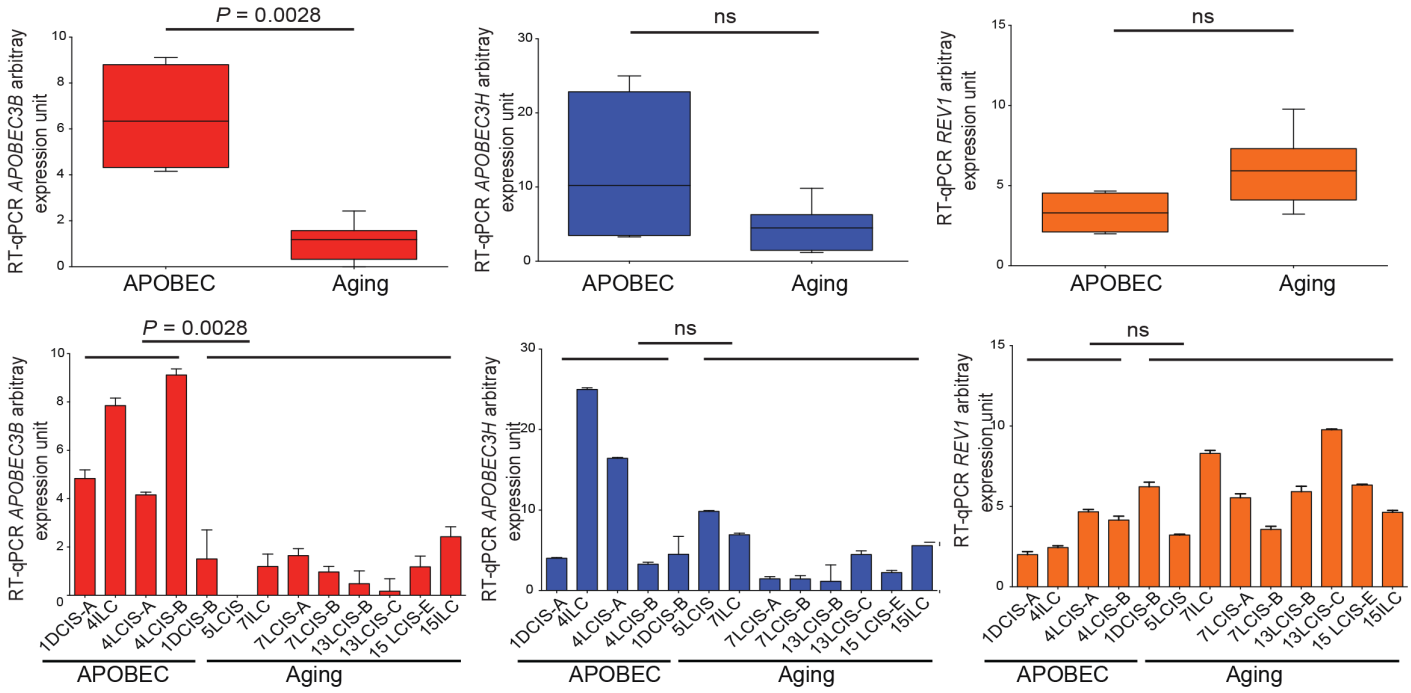
B



C



D



non significant (ns) p-value > 0.05