SUPPLEMENTARY METHODS

Subjects and samples

Following pathology review, 27 ductal carcinoma *in situ* (DCIS) lesions synchronously diagnosed with invasive ductal carcinoma of no special type (IDC-NST; n=26), including two cases of multifocal/ multicentric DCIS, and 7 DCIS not associated with invasion (i.e., pure DCIS) were included in this study (**Table 1**). IDC-NSTs were graded according to the Nottingham grading system (1) and the nuclear grading of DCIS was performed following the recommendations by the College of American Pathologists (2). Estrogen receptor (ER) and HER2 status were evaluated by immunohistochemistry and/or fluorescence *in situ* hybridization (FISH) by five pathologists (FGP, RB, FCG and MV and HYW) according to the American Society of Clinical Oncology (ASCO)/ College of American Pathologists (CAP) guidelines (3,4), as previously described (5).

Whole exome sequencing and MSK-IMPACT sequencing

In brief, after aligning reads to the reference human genome GRCh37, somatic mutations 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), mutations identified in at least one sample were subsequently interrogated in all related samples of a given patient using SAMtools mpileup (version 1.2 htslib 1.2.1) (6). The potential functional effect of somatic mutations was defined using a combination of predictors with a high negative predictive value (7), as previously described (8), and genes were annotated according to their presence in three cancer gene datasets, Bailey *et al*. (9), the Cancer Gene Census (10) and Lawrence *et al*. (11). Mutations affecting hotspot codons were annotated according to Chang et al (12), as previously described (13,14). Allele-specific copy number alterations (CNAs) and loss of heterozygosity (LOH) for specific genes were defined using FACETS (15), as previously described (8,14,16), and purity and ploidy estimations were calculated using ABSOLUTE (17).

Targeted amplicon re-sequencing validation of somatic mutations

Validation of the mutations detected by whole exome sequencing (WES) was performed for cases with sufficient DNA (n=10), using a custom designed AmpliSeq panel. Out of 3,694 somatic mutations identified by WES or MSK-IMPACT, 652 were investigated in 12 DCIS and 10 IDCs from cases 2, 4-10, 12, and 13. Of the mutations tested, 617 (95%) 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 downstream analyses, as previously described (8). Given the high accuracy of the mutation detection based on the pipeline employed for WES and MSK-IMPACT analysis, the mutations not subjected to validation were included in the downstream analyses.

Clonal frequencies

The mutant allelic fraction measurements were transformed into estimates of clonal frequencies jointly for all lesions from a given patient using a Dirichlet clustering model, which simultaneously estimates the genotype and clonal frequency given a list of somatic mutations and their local copy number. Purity and ploidy estimates, as well as modal copy number from ABSOLUTE (17) were employed as the input data for PyClone.

Truncal and branch mutations

For all cases of clonally-related DCIS and IDC-NSTs, mutations were categorized as truncal or branch using PyClone (18). Truncal mutations were defined as those concurrently present in the modal populations of all DCIS lesions and IDC-NSTs from a given patient. All non-truncal mutations were defined as branch mutations.

Measures of diversity

The Shannon index is borrowed from information theory and summarizes the diversity of a population in a number. It is defined as H =$-\sum_{i=1}^{n} p_i \mathrm{x} \ln(p_i)$, where H is the Shannon index metric, p_i is the percentage of a subpopulation in the overall population and n is the number of subpopulations. The Gini-Simpson index is defined as the probability that two entities taken randomly from the dataset of interest represent different types and is defined as D = $1-\sum_{i=1}^{n} {p_i}^2$ where D is the Gini-Simpson index metric, p_i is the percentage of a subpopulation in the overall population and n is the number of subpopulations. In this study, for both the Shannon and Gini-Simpson indices, p_i and n were defined as the percentage of a genetically distinct subclone within a lesion and the number of subclones, respectively, derived from the tumor clone structure inferred using Pyclone (18), as previously described (19).

Phylogenetic tree construction

Maximum parsimony trees were inferred using binary presence/absence matrices built from somatic genetic alterations, including synonymous and non-synonymous SNVs, indels, within the clonally-related lesions from each patient as described in Murugaesu et al. (8,20). For the construction of phylogenetic trees based on CNAs, major and minor copy numbers computed by FACETS (15) were modeled using transducer-based pairwise comparison functions using the program MEDICC (21) assuming a diploid outgroup with no CNAs to root the phylogenies. Only regions with a total copy number ≤8 were included in this analysis to increase the accuracy of phasing into parental copy number states. Support values for the phylogenetic trees were obtained by resampling the pairwise distance matrix 100 times with added Gaussian noise and by counting similar bipartitions between the resulting trees and the original phylogeny.

Comparisons with invasive breast cancers from The Cancer Genome Atlas (TCGA)

For comparisons with the TCGA dataset, clinicopathologic data were retrieved from Riaz et al (22). The publicly available MC3 dataset was retrieved from the TCGA Pan-Cancer Analysis (23) at https://gdc.cancer.gov/about-data/publications/mc3-2017. Previous studies have demonstrated the equivalence between the TCGA MC3 dataset and the pipeline employed in this study for mutation detection (14,19).

SUPPLEMENTARY REFERENCES

1. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology **1991**;19(5):403-10.
2. Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL*, et al.* Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. Arch Pathol Lab Med **2009**;133(1):15-25.
3. Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S*, et al.* American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. J Clin Oncol **2010**;28(16):2784-95.
4. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS*, et al.* Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. J Clin Oncol **2018**;36(20):2105-22.
5. 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(3):376-85.
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. Bioinformatics **2009**;25(16):2078-9.

7. 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.

8. 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.

9. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A*, et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell **2018**;174(4):1034-5.

10. 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.

11. 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.

12. Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P*, et al.* Accelerating Discovery of Functional Mutant Alleles in Cancer. Cancer Discov **2018**;8(2):174-83.

13. Geyer FC, Li A, Papanastasiou AD, Smith A, Selenica P, Burke KA*, et al.* Recurrent hotspot mutations in HRAS Q61 and PI3K-AKT pathway genes as drivers of breast adenomyoepitheliomas. Nat Commun **2018**;9(1):1816.

14. 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**;110(9):1030-4.

15. 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.

16. Pareja F, Brandes AH, Basili T, Selenica P, Geyer FC, Fan D*, et al.* Loss-of-function mutations in ATP6AP1 and ATP6AP2 in granular cell tumors. Nat Commun **2018**;9(1):3533.

17. 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.

18. 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.

19. Lee JY, Schizas M, Geyer FC, Selenica P, Piscuoglio S, Sakr RA*, et al.* Lobular carcinomas in situ display intra-lesion genetic heterogeneity and clonal evolution in the progression to invasive lobular carcinoma. Clin Cancer Res **2018**;25(2):674-86.

20. 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**;5(8):821-31.

21. Schwarz RF, Trinh A, Sipos B, Brenton JD, Goldman N, Markowetz F. Phylogenetic quantification of intra-tumour heterogeneity. PLoS Comput Biol **2014**;10(4):e1003535.

22. Riaz N, Blecua P, Lim RS, Shen R, Higginson DS, Weinhold N*, et al.* Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. Nat Commun **2017**;8(1):857.

23. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C*, et al.* Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst **2018**;6(3):271-81 e7.