# Timing evolution of lobular breast cancer through phylogenetic analysis

Danai Fimereli,[a,1] David Venet,[a,1] Mattia Rediti,[a] Bram Boeckx,[b,c] Marion Maetens,[a,d] Samira Majjaj,[a] Ghizlane Rouas,[a] Caterina Marchio,[e,f] Francois Bertucci,[g] Odette Mariani,[h] Maria Capra,[i] Giuseppina Bonizzi,[i] Federica Contaldo,[i] Christine Galant,[j,k] Gert Van den Eynden,[l] Roberto Salgado,[m,n] Elia Biganzoli,[o] Anne Vincent-Salomon,[h] Giancarlo Pruneri,[p,q] Denis Larsimont,[r] Diether Lambrechts,[b,c] Christine Desmedt,[d] David N. Brown,[a,s] Françoise Rothé,[a,1] and Christos Sotiriou [a,1]*

[a]J.-C. Heuson Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium
[b]Laboratory of Translational Genetics, VIB Center for Cancer Biology, Leuven, Belgium
[c]Laboratory of Translational Genetics, Department of Human Genetics, KU Leuven, Leuven, Belgium
[d]Laboratory for Translational Breast Cancer Research, Department of Oncology, KU Leuven, Leuven, Belgium
[e]Department of Medical Sciences, University of Turin, Turin, Italy
[f]FPO-IRCCS Candiolo Cancer Institute, Candiolo, Italy
[g]Predictive Oncology Laboratory, Institut Paoli-Calmettes, CRCM, INSERM U1068, CNRS UMR7258, Aix-Marseille Université Marseille, France
[h]Department of Pathology, Institut Curie, Paris Sciences Lettres Research University, Paris, France
[i]Biobank for Translational and Digital Medicine, IEO, European Institute of Oncology IRCCS, Milan, Italy
[j]Department of Pathology, Cliniques Universitaires Saint Luc, Brussels, Belgium
[k]IREC, Université Catholique de Louvain, Brussels, Belgium
[l]Department of Pathology, Sint Augustinus, Wilrijk, Belgium
[m]Department of Pathology, GZA-ZNA Hospitals, Antwerp, Belgium
[n]Division of Research, Peter Mac Callum Cancer Centre, Melbourne, Australia
[o]Department of Biomedical and Clinical Sciences (DIBIC) "L. Sacco" & DSRC, LITA Vialba campus, University of Milan, Milan, Italy
[p]Division of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy
[q]School of Medicine, University of Milan, Milano, Milan, Italy
[r]Department of Pathology, Institut Jules Bordet, Brussels, Belgium
[s]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

## Summary

**Background** Late distant recurrence is a challenge for the treatment of invasive lobular carcinoma (ILC) of the breast. Despite in-depth characterisation of primary ILC, the molecular landscape of metastatic ILC is still only partially understood.

**Methods** We retrospectively identified 38 ILC patients from the tissue banks of six European institutions. DNA extracted from patient matched primary and metastatic FFPE tissue blocks was whole genome sequenced to compute somatic copy number aberrations. This, in turn, was used to infer the evolutionary history of these patients.

**Findings** The data show different metastatic seeding patterns, with both an early and late divergence of the metastatic lineage observed in ILC. Additionally, cascading dissemination from a metastatic precursor was a dominant rule. Alterations in key cancer driver genes, such as *TP53* or *CCND1*, were acquired early while additional aberrations were present only in the metastatic branch. In about 30% of the patients, the metastatic lineage harboured less aberrations than the primary tumour suggesting a period of tumour dormancy or prolonged adaptation at the distant site. This phenomenon was mostly observed in tumours from de novo metastatic patients.

**Interpretation** Our results provide insights into ILC evolution and offer potential paths for optimised ILC care.

**eBioMedicine 2022;82: 104169**
Published online 23 July 2022
https://doi.org/10.1016/j.ebiom.2022.104169

*Corresponding author at: Breast Cancer Translational Research Laboratory (BCTL), Université Libre de Bruxelles, Institut Jules Bordet, Rue Meylemeersh 90, 1070 Anderlecht, Belgium.
*E-mail address:* christos.sotiriou@bordet.be (C. Sotiriou).

[1] These authors contributed equally.

**Funding** This work has received financial support from Les Amis de l'Institut Bordet, MEDIC, the Breast Cancer Research Foundation (BCRF) and the Belgian Fonds National de la Recherche Scientifique (F.R.S-FNRS).



**Keywords:** Breast cancer; Distant metastasis; Heterogeneity; Metastatic dissemination; Tumour progression

### Research in context

*Evidence before this study*

Invasive lobular carcinoma (ILC) of the breast is the most common special type of breast cancer. Despite therapeutic advancements, a substantial number of patients with treated early-stage ILC will still relapse as late as 20 years after the initial tumour diagnosis, thus understanding the mechanisms by which genomic aberrations shape tumour evolution from the primary to the metastatic site is of crucial importance. However, the characterization and timing of the genomic events that best portray the trajectory of cancer cells from the primary tumour to metastasis in ILC is still largely lacking.

*Added value of this study*

In this study, we employed whole genome sequencing to unravel the underlying genomic aberrations and to explore tumour evolution through the reconstruction of phylogenetic trees. We demonstrate that different metastatic seeding patterns can occur in ILC, with both early and late dissemination events possible. Genomic aberrations in key driver genes were detected early during tumour evolution, however acquired new driver aberrations were also observed. Notably, in a subset of patients, the metastasis harbored less genomic aberrations than the primary tumour, indicating that possibly a decelerated tumour progression is in place, associating with de novo metastatic disease.

*Implications of all the available evidence*

Collectively, our study provided an understanding of how ILC evolves through time, with potential implications for the clinical management of ILC.

## Introduction

Recurrence in distant organs is the major cause of cancer-related deaths despite significant progress in the understanding and the treatment of primary cancer. Approximately 20-30% of patients treated for early-stage breast cancer (BC) still develop metastatic disease[1] whereby the route of metastatic dissemination can differ even within this single cancer type. In fact, previous work from our group and others have shown that primary BC can spread through the lymphatic or the hematogenous route.[2,3] Adding another layer of complexity, the time at which overt distant recurrences manifest is variable ranging from a few months up to twenty years after primary tumour diagnosis and treatment. For example, despite the therapeutic advancement leading to lower rates of distant metastasis and mortality for patients with hormone receptor-positive BC treated with adjuvant endocrine therapy, many patients will still suffer late distant recurrence, even 10 or 20 years later.[4,5]

Invasive lobular carcinoma (ILC) of the breast is the most common "special" type of breast cancer, after invasive BC "of no special type" (known as invasive ductal carcinoma) and accounts for approximately 5-15% of all cases of invasive BC.[6] It is defined by distinct biological, morphological, and clinical characteristics like low histological grade and low proliferation rate; also, through the expression of estrogen receptor (ER) (encoded by the *ESR1* gene), lack of *HER2/neu* (*ERBB2*) amplification, as well as typically the lack of E-cadherin (encoded by the *CDH1* gene) which can occur through various mechanisms that involve mutation, deletion and/or methylation. A number of somatic alterations, including mutations in genes like *ESR1, FOXA1, GATA3* or genes from the PI3K pathway and large-scale aberrations including recurrent gains, amplifications or deletions in 1q, 8q, 11q13 and 16q among other, paint the genome-wide landscape of early ILC.[7,8] Despite the in-depth characterisation of early ILC, the molecular landscape of metastatic ILC has been only recently but yet still partially portrayed, unravelling a number of acquired mutations in key genes like *ESR1, AKT1, NF1, MAP3K1* as well as copy number aberrations including *CCND1, CCNE1* and *IGF1R* amplifications.[9–12] Some of these mutations, associated to endocrine resistance, could provide a link and an explanation to the late recurrence observed in ER-positive ILC. However, acknowledging the difficulties of obtaining matched primary tumours and distant metastases from the same patient, characterisation of the timing of metastatic dissemination and the intratumour heterogeneity of metastatic ILC is still widely lacking.

Here we analysed somatic copy number aberrations obtained from whole genome sequencing data of matched primary tumour and metastases from 38 patients with ILC. We inferred the evolutionary histories of these ILC patients through phylogenetic reconstructions and assessed driver gene heterogeneity between primary and

metastatic samples. We observed different patterns of metastatic seeding and described occurrences of a decelerated tumour progression or tumour dormancy, providing further insight into ILC evolution.

## Methods

### Patients' selection and samples' collection

We retrospectively identified a cohort of breast cancer patients from the tissue banks of the Institut Jules Bordet (Belgium), the Cliniques Universitaires Saint Luc (Belgium), GZA Ziekenhuizen (Belgium), the European Institute of Oncology (Italy), the Institut Paoli-Calmettes (France) and the Institut Curie (France).[12,9] Inclusion criteria were patients with primary lobular breast tumour and a confirmed distant organ metastasis of breast cancer origin as per medical records. Eligible patients included only those for whom at least one tumour and one distant metastasis, as well as a histologically normal tissue sample as germline reference were available as formalin-fixed paraffin-embedded tissue blocks.

### Ethics

The project has been approved by the ethics committee of the Institute Jules Bordet (N°2504). Given the retrospective nature of the study, the ethics committee did not require the patients to sign an informed consent.

### Histopathological characterisation

Histopathological evaluation of the FFPE tissue samples involved histological subtyping, assessment of histological grade, and determination of the percentage of tumour epithelial cells. Samples with minimum tumour cellularity of 20% (if <20%, then only considered if microdissection could be done) were kept. Immunohistochemical evaluations of estrogen receptor, progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2) and Ki67 were retrieved from local routine assessments. Stromal tumour infiltrating lymphocytes (sTILs) were assessed by two experienced BC pathologists (R.S and G V.D.E.) following the international guidelines.[13] The average of the two scores was considered as the final value for that sample. De novo metastatic patients were defined as those patients with metastatic BC at initial diagnosis.

### DNA extraction and whole genome sequencing

DNA extracted from the tumour and normal matched FFPE samples were sequenced at low pass whole genome coverage in collaboration with the Vlaams Instituut voor Biotechnologie (Leuven, Belgium). Whole genome shotgun libraries were prepared using the KAPA library preparation kit according to the manufacturer's instructions from 100ng of double stranded DNA. After quantification by qPCR, the resulting libraries were sequenced on an Illumina HiSeq 4000 series device in 51bp single-end mode.

### Low pass whole genome sequencing and phylogenetic tree reconstruction

A total of 99 samples from 38 patients were sequenced to an average target coverage of 0.5X. The raw sequence reads were aligned to the human genome reference hg19/GRCh37 using the BWA aligner[14] and duplicate reads were marked using Picard.[15] In order to infer $\log_2$ ratio estimates of copy numbers from the depth of sequence coverage, the sequenced reads from the aligned and sorted BAM files were binned into equally spaced 15kbp windows and corrected for library size and GC content using CNVkit.[16] Sequencing-inaccessible regions like centromere and telomeres were removed. The matched normal samples were used to compute a reference calibration set and the corrected read depths from the tumour samples were then transformed into $\log_2$ ratio estimates by reference to the pool of normal matched samples. The genome-wide $\log_2$ ratio profiles were used to compute the median absolute pairwise deviation (MAPD) and median auto-correlation (MAC) as quality control metrics whereby a MAPD>0.3 or a MAC>0.5 were used as thresholds to flag low quality samples. The samples passing this quality control were then segmented using the multitrack penalized least square regression method of Nilsen et al.[17] whereby all samples belonging to a given patient were processed simultaneously to define common breakpoints. The segmented $\log_2$ ratios were further used as input to ABSOLUTE[18] in order to infer the cancer cell fraction (CCF), genomic mass, and segment wise copy numbers. As final quality control, all samples with a CCF<0.1 were discarded from downstream analyses.

To infer the phylogenetic trees, we obtained the continuous estimates of copy numbers and rounded them to the nearest integer value. These values were used as input to CNT-ILP,[19] that was run with default parameter values. For tree reconstruction, a pure diploid outgroup with no copy number aberrations at any loci is assumed for rooting the phylogenies and reconstructing the ancestral states. Statistical support for the phylogenetic trees was computed through bootstrapping. For each patient, the $N$-by-$n$ matrix of integer copy numbers, where $N$ is the number of samples and $n$ is the number of genomic loci was resampled with replacement along the $n$ columns to create 50 similar sized matrices which were used as input for phylogenetic reconstruction. Bootstrap percentage values correspond to the number of similar bipartitions between the results trees and the original phylogeny.

### Calculation of genetic distance

For analysing the genetic distances between primary tumour and metastasis, in the subset of patients with multiple primary/metastasis samples, we preserved only one (1) primary-metastasis pair and recalculated the phylogenetic trees. During this step, three patients were removed from further analysis, due to large

differences in the copy number profiles between the matched primary and metastatic samples. For the remaining patients, the representative pair was chosen by applying the following rules: 1) In the case of multiple metastatic samples clustering together but in a separate branch from primary samples, we kept the metastatic sample with the highest purity; 2) In the case of multiple primary samples clustered together in a separate branch from metastatic samples, we kept the primary sample with the highest purity; 3) In the case of a branch containing a primary-metastasis pair and clustered separately from other primary samples, we kept that pair. For patient PT02, where the two metastatic samples come from two different sites biopsied at different time points, we kept the bone metastasis as it was obtained closer in time from the primary tumour sample. Distances were calculated from the phylogenetic tree as illustrated in Figure 3a. Specifically, *dTrunk* was calculated as the number of events common to primary and metastasis, from the root until the most recent common ancestor; *dPrimary* and *dMeta* were calculated as the number of events starting from the most recent common ancestor until primary/metastasis sample tip. The normalised genetic distance was calculated as:

$$Genetic\ distance = (dMeta - dPrimary)/(dPrimary/2 + dMeta/2 + dTrunk)$$

Based on the genetic distance, we categorised patients into three classes: 1) "positive" class with genetic distance ≥ 0.03; 2) "negative" class with genetic distance < -0.03; 3) "neutral" class with -0.03< *genetic distance* <0.03. Ploidy corrected copy number values were used to call copy number changes with the following thresholds: gain, between 2.5 and 4 copies; amplification >4 copies, loss <1.6 copies and deep deletion <0.7 copies. Potential driver genes were obtained from Nik-Zainal et al.[20]

The trunk to primary ratio was calculated as:

$$ratio\ trunk/primary = dTrunk/(dPrimary + dTrunk)$$

### Statistical analysis

Comparisons between categorical variables were performed using Fisher's exact test while comparisons between continuous variables were performed using Mann-Whitney *U* test. All tests were two-sided and p-values<0.05 were considered significant. The p-values generated are descriptive and were not adjusted for multiple testing. All statistical analyses and plots (including plotting of the trees) were done in R v4.0.2. Heatmaps were plotted with R package ComplexHeatmap.[21]

### Role of funding source

The funders had no role in study design, data collection, data analysis, interpretation or writing of the report.

## Results

### Patients and tumour characteristics

In order to investigate the evolutionary trajectories of metastatic ILC, we retrospectively selected 38 patients with matched primary and metastatic samples from 6 European institutes, as previously described.[12,9] Low pass whole genome sequencing was performed on 99 samples and the calculated copy number aberrations (CNA) were used to draw phylogenetic trees for each patient (Figure 1a). In total, 14 patients had multiple primary and/or metastatic samples available while 24 patients had a single sample for the primary and for the matched metastatic lesion available. The primary tumour samples ranged between 1-4 per patient while the metastatic samples ranged from 1 to 3 per patient. The most common metastatic sites were bone ($n$ = 11), digestive tract ($n$ = 7), skin ($n$ = 6), pleura ($n$ = 5) and liver ($n$ = 4) followed by other sites (Supplementary Figure S2). In brief, in our cohort 58% of patients were postmenopausal, 16% were de novo metastatic and 76% had a tumour size more than 2cm. All primary tumours were ER-positive, 74% ($N$ = 28/38) were PgR-positive, 5% ($N$ = 2/38) were HER2-positive and 50% ($N$ = 19/38) were of classical histological subtype. Additionally, 95% ($N$ = 36/38) of patients had received adjuvant endocrine therapy and 60% ($N$ = 23/38) had received adjuvant chemotherapy. The detailed clinicopathological characteristics of the cohort are presented in Table 1 and Table S1.

### Phylogenetic reconstruction of metastatic ILC

The somatic copy number profiles of all patients were used to obtain the phylogenetic trees (Supplementary Figure S4). Detailed phylogenies were reconstructed for patients with multiple matched primary and metastatic samples, thus allowing us to explore the routes of metastatic dissemination. Based upon the clustering patterns, we observed that in 9/14 patients (PT02, PT04, PT07, PT14, PT42, PT43, PT45, PT63 and PT64), the metastatic sample(s) clustered separately from the primary samples indicating that the metastatic lineage diverged early, at the point of the most recent common ancestor. In the examples of patients PT07 and PT42, all primary samples clustered together and separately from the metastatic sample (bone and pleura respectively) that continued to evolve and accumulate mutations (Figure 1b-c). In parallel, in 5/14 patients (PT06, PT36, PT55, PT58 and PT67) the metastatic sample diverged at a later evolutionary stage during diversification of the primary tumour. In the example of patient PT06 who was de novo metastatic, the biopsied lesion in bone was more genetically related to one region of the primary tumour and thus clustered together, while the additional primary samples evolved in a separate branch. Similarly, in patient PT36 who was de novo metastatic, we observed a co-clustering of one primary

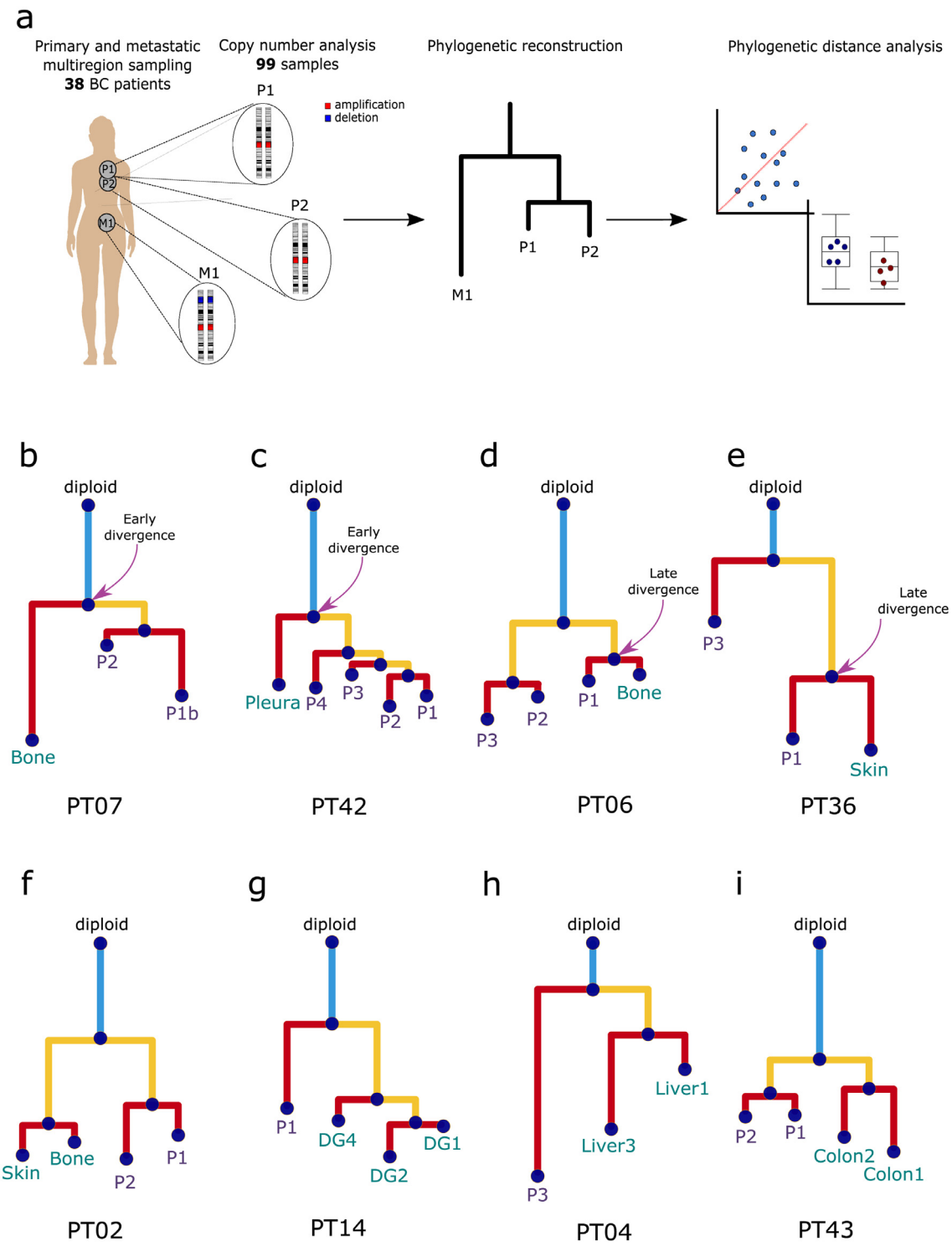


**Figure 1. Phylogenetic reconstruction of lobular breast cancer genomes.** a) schematic representation of the design study. Primary tumours (P) and metastatic samples (M) were obtained from 38 patients with lobular breast cancer. b-i) Phylogenetic trees of representative patients.

| | All patients *N* = 38 |
|---|---|
| Age | |
| <50 years | 9 (23.7%) |
| >=50 years | 29 (76.3%) |
| Menopause status | |
| Pre/Peri | 16 (42.1%) |
| Post | 22 (57.9%) |
| De novo metastatic | |
| Yes | 6 (15.8%) |
| No | 32 (84.2%) |
| Primary tumour size | |
| <2cm | 9 (24.3%) |
| ≥2cm | 28 (75.7%) |
| Missing | 1 |
| Histological subtype | |
| Classic | 19 (50.0%) |
| Non-Classic | 18 (47.4%) |
| Missing | 1 (2.6%) |
| PgR status (primary) | |
| Negative | 10 (26.3%) |
| Positive | 28 (73.7%) |
| HER2 status (primary) | |
| Negative | 34 (89.5%) |
| Positive | 2 (5.3%) |
| Missing | 2 (5.3%) |
| Tumour grade (primary) | |
| 1 | 7 (18.4%) |
| 2 | 24 (63.2%) |
| 3 | 7 (18.4%) |
| Nodal status | |
| Negative | 16 (42.1%) |
| Positive | 21 (55.3%) |
| Missing | 1 (2.6%) |
| Adjuvant endocrine therapy | |
| Yes | 36 (95%) |
| No | 2 (5%) |
| Adjuvant chemotherapy | |
| Yes | 23 (60.5%) |
| No | 15 (39.5%) |

***Table 1*: Clinical features of the patient cohort.**

sample with the metastasis (skin lesion), separately from the second primary sample (Figure 1d-e). Together, these results show that both early and late divergence of the metastatic lineage are possible in ILC, associated with an underlying intratumour heterogeneity.[22] Focusing on the four patients with multiple metastatic samples sequenced (PT02, PT04, PT14 and PT43), it was observed that in all cases, the metastases were more closely related to each other than to the primary tumour (Figure 1f-i). Patient PT02 had metachronous tumours in bone and skin biopsied at different timepoints. The phylogenetic analysis revealed that the two metastases clustered together and separately from the primary tumour samples. Similarly, in patient PT14, the primary tumour was an outgroup to all the metastases, while the metastatic samples biopsied from different lesions at different timepoints (two localised in the small intestine, one in the colon) clustered together. These data could point towards different explanations for the phylogenetic similarity of the metastases in our cohort. One possibility is that these metastases were seeded independently by a primary tumour clone with high metastatic capacity.[23] An alternative explanation could point out that these metastases were seeded from a common metastatic precursor indicating a cascading dissemination, as previously described.[24]

To further assess the underlying heterogeneity, we analysed CNAs in a set of known driver genes[20] by using the matched primary and metastatic samples. Gene change (loss or gain) was called if the corresponding gene copy number is different from the sample's ploidy at different copy levels. In the case of patients with multiple primary and/or metastatic samples, the change was defined as present if the gene was altered in at least one sample (see details in Material and Methods). As shown in Figure 2, multiple driver genes were located in the trunk of the tree, which is being shared by both the primary tumour and metastasis. The most frequent alterations include deletions in *CDH1, TP53, RB1* and *MAP2K4* (92%, 47%, 39% and 37% of patients respectively) and gains in *MYC, CCND1, ZNF217* and *MDM2* (32%, 21%, 21% and 16% of patients respectively). In parallel, alterations private to the metastasis (that are not altered in the matching primary sample) were also observed, with the most frequent involving deletions in *FGFR2, PTEN* and *NCOR1* (21%, 18%, and 18% of patients respectively) and gains in *AKT2, ZNF217, CCND1* and *EGFR* (16%, 16%, 13% and 11% of patients respectively). These results confirm the existing evidence that aberrations in key driver genes are present early during tumourigenesis but also that metastasis continues to acquire aberrations even after dissemination.

### Evidence of decelerated metastatic progression

While evaluating the sample trees of the whole cohort, we observed that, for a subset of patients, the branch length of the metastatic lineage was shorter than the branch of the primary lineage suggesting a decelerated metastatic progression. To better estimate the prevalence of this event, we first computed for each patient a genetic distance defined as the difference in the number of aberrations between the metastatic and the primary samples normalised by the total tree length (Figure 3a). While such a quantification is straightforward for patients with a single primary/metastasis sample pair, for those patients with multiple samples we only kept one representative primary-metastasis pair for the analysis (see details in Material and Methods). Patients were

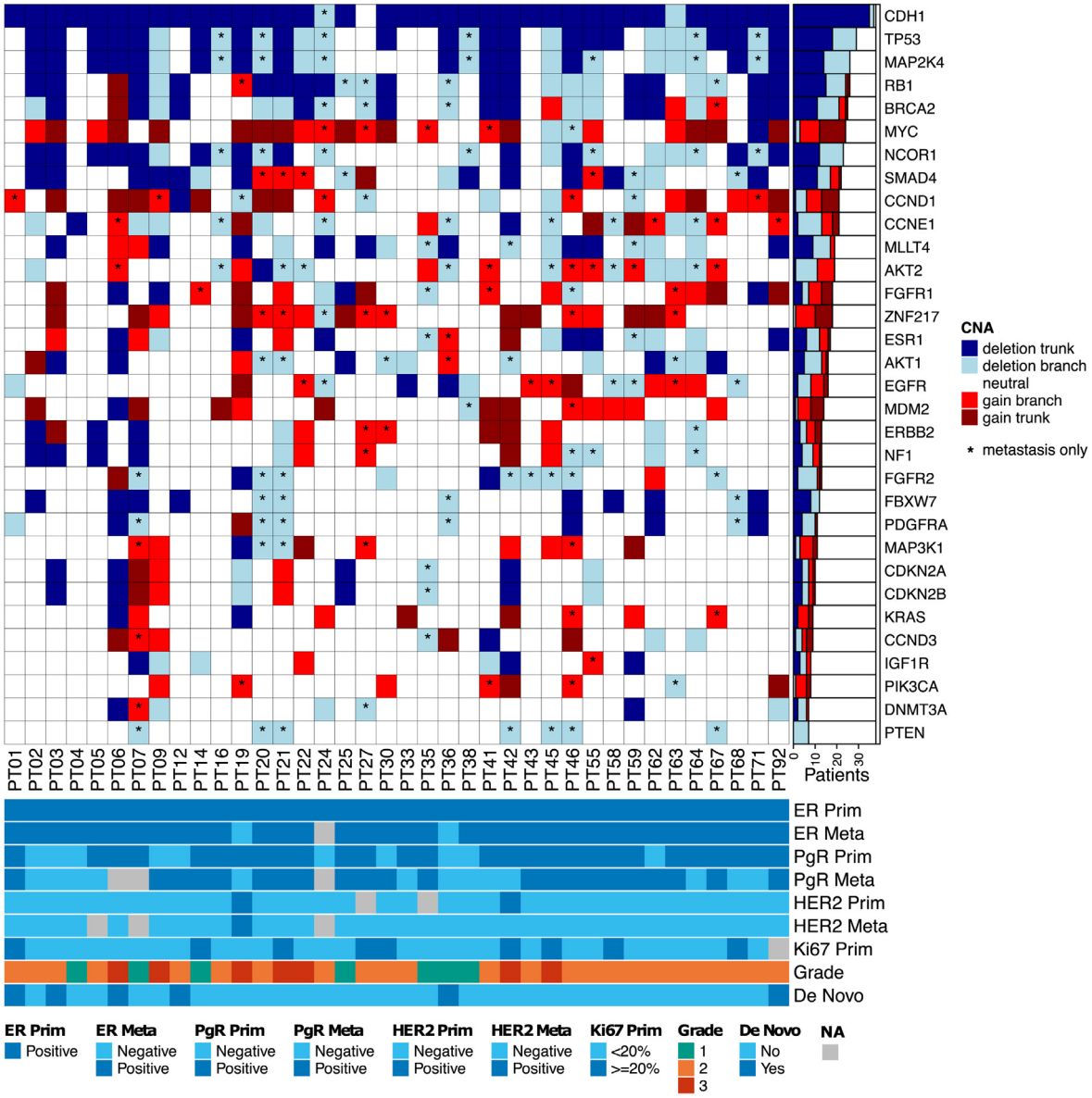


**Figure 2. Distribution of copy number aberrations across 32 driver genes.** The heatmap represents the status of copy number aberrations in the trunk of the tree or in one of the branches (primary or metastasis). The asterisk indicates the presence of the aberration only in the metastatic sample of that patient. Estrogen receptor (ER), progesterone receptor (PgR) and HER2 in the primary (Prim) tumour and metastasis (Meta).

then categorised based on their genetic distances into three classes capturing distinct dynamics of metastatic progression (Figure 3a). As illustrated in Figure 3b, we observed a higher number of aberrations in the metastasis as compared to the primary after diverging from the common ancestor defined as the "positive" class in 20 out of 35 (57%) informative patients. Unexpectedly, for 11 out of 35 (31%) patients the metastatic branch was shorter than the primary branch, indicating that the metastasis accumulated less aberrations that the matched primary tumour ("negative" class). A small subset of four patients (12%) had similar number of aberrations in both the primary and the metastatic samples defined as the "neutral" class. Analysing the branch lengths of the primary and metastatic lineages, we observed a significant correlation between primary and metastatic samples (Spearman's correlation rho=0.66, p-value=1.7e-05, Supplementary Figure S5). Lastly, the genetic distance only loosely correlated with the time passed between time of diagnosis of the primary tumour and metastasis biopsy (Spearman's correlation rho=0.23, Figure 3c).

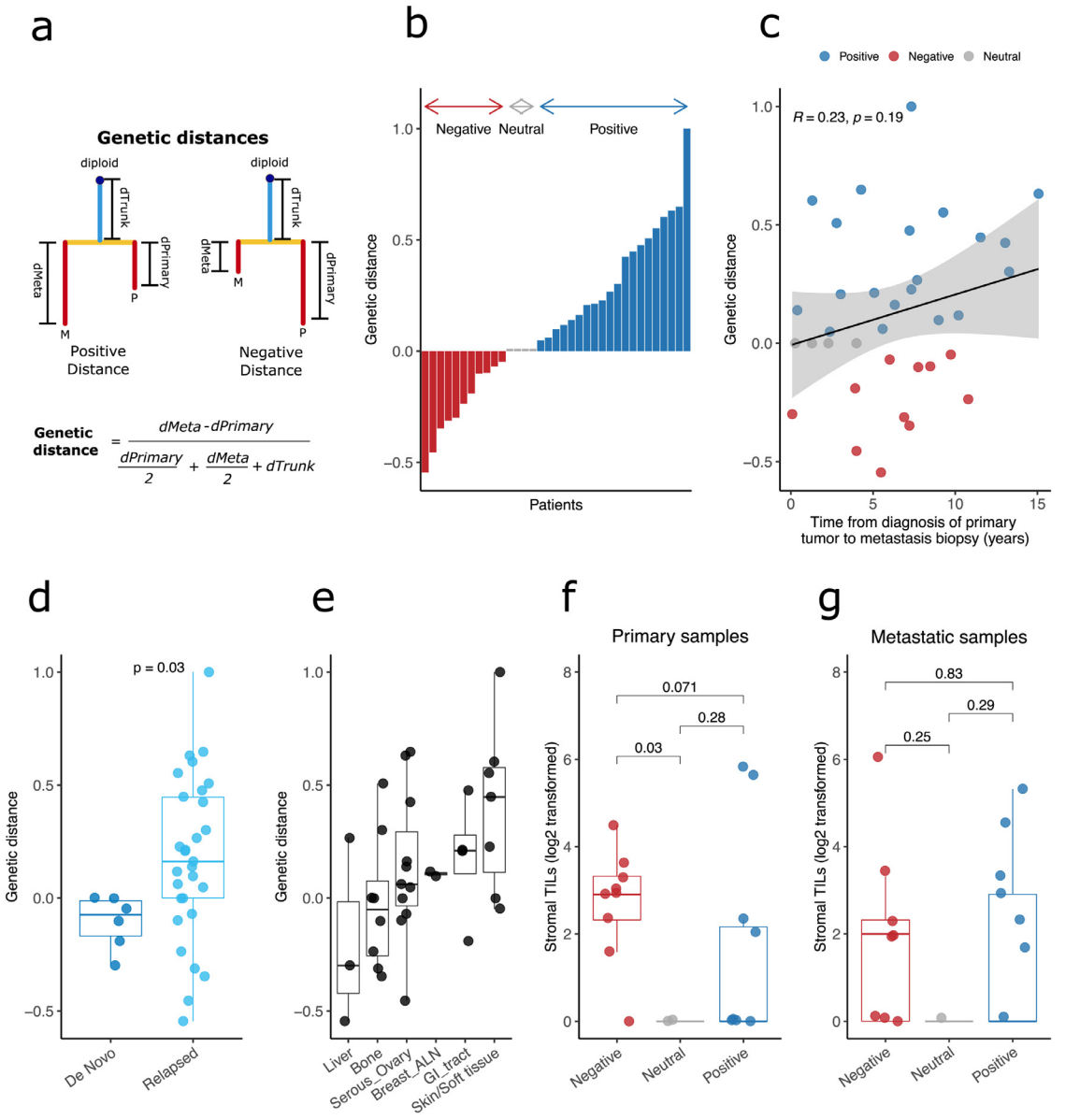


**Figure 3. Presence of positive and negative genetic distances in lobular breast cancer.**) Phylogenetic model trees indicating examples of a positive and a negative genetic distance. Genetic distance is defined as by the equation. b) Distribution of genetic distances in the cohort ($n$ = 35 informative patients). Each bar represents a patient. c) Genetic distance associated with the time from diagnosis of primary tumour to metastatic biopsy (Spearman's rank correlation). d) Boxplot indicating the genetic distance in de novo metastatic ($n$ = 6) and later relapsed patients ($n$ = 29). e) Boxplots indicating the genetic distance at different metastatic sites. f-g) Boxplots indicating the distribution of stromal TILs across the three classes of genetic distances, in primary and metastatic samples respectively. In d-g each dot represents a patient and P values were calculated using a two-sided Mann-Whitney $U$ test.

### Clinicopathological and molecular characterisation of genetic distance

Next, we compared the clinicopathological characteristics of our cohort between the two opposing classes ("positive" / "negative"). No difference/association was observed with age, menopausal status, tumour size, grade, PgR/HER2 status, nodal status, number of positive nodes or tumour histology regardless of using the genetic distance as a continuous or as a dichotomous variable. Interestingly, an association between the two classes and de novo metastatic status was observed, with the majority (4/6) of the de novo patients being assigned to the "negative" class (p-value=0.01, Fisher's exact test) and showing lower distance compared to the rest of the relapsed patients (p-value=0.03, Mann-Whitney $U$ test, Figure 3d) despite the fact that the metastatic samples in

the de novo metastatic patients were taken later in the history of the patient. In addition, patients metastasizing to different sites showed varying distances, with liver and bone metastases having the lowest distance and skin the highest, albeit without significant differences between the sites (Figure 3e). Finally, a trend highlighting higher levels of stromal TILs in the primary tumour was observed in the "negative" class compared to the "positive" class (Figure 3f-g). At the genomic level, comparing the copy number aberrations of the 32 driver genes in the primary only samples between the two genetic distance classes revealed that *NF1, AKT1* and *FGFR2* were more frequently altered in the "negative" class compared to the "positive" class (p-value=0.04, 0.01 and 0.03 respectively, Fisher's exact test, Supplementary Figure S6). Additionally, the *AKT1* gene was also more often deleted in the trunk of the phylogenetic tree in the "negative" class (p-value=0.04, Fisher's exact test). On the contrary, in the metastatic samples, no driver gene was significantly different between the two classes. Taken together, these data could denote a decelerated tumour progression for at least a subset of patients, either in the form of slower cell cycle or possibly dormant disseminated tumour cells. The absence of common copy number aberrations in these patients could also suggest that the tumour microenvironment or other nongenetic processes play an important role.

Lastly, using the single-pair phylogenetic trees, we computed the ratio of the common trunk branch length relative to the sum of the common trunk and the primary branch length, as to represent the number of aberrations that accumulate after the primary and metastatic lineages divergence. As seen from our data (Figure S1a), in 22/35 (63%) patients, the bulk of the aberrations accumulated in the trunk of the phylogenetic tree thus having long trunks and short primary branches whereas in 8/35 (23%) patients the length of the trunks was shorter compared to the primary branches (5/35 patients were classified as intermediate cases). No differential copy number change in specific driver genes between these two categories was observed. Although no association was observed with clinicopathological characteristics or the genetic distance, we found a loose, albeit not significant, inverse correlation with relapse time (Spearman's correlation rho=-0.28, Figure S1b). These results show that in ILC, the majority of CNAs are acquired early during tumour evolution.

## Discussion

In this study, we described the evolutionary trajectories and the intratumour heterogeneity of ILC though the analysis of low pass whole genome sequencing data from matched primary and metastatic samples. The delineation of the phylogenetic trees unraveled different mechanisms of metastatic seeding. Namely, in 64% of the patients, we observed a separate clustering of the metastatic and the primary samples, indicating an early genetic diverging point. On the other hand, in 36% of the patients, the metastatic sample was more genetically related to a specific primary sample than the rest of the primary samples, indicating that a specific parental primary clone seeded the metastasis. Furthermore, we observed that in the cases of patients with multiple metastases available, all the metastases clustered together, suggesting seeding from a common "metastatic precursor". These results are in line with previous reports indicating heterogeneity within multiple regions of the same lesion and highlight the necessity of multiregion genetic analysis of tumours.[25] Such multiregion analysis employed by previous studies in breast cancer showed indeed that different scenarios of metastatic seeding exist, with either a monoclonal or a polyclonal origin of metastasis.[24,3] Irrespective of the dissemination route, known driver genes including *TP53, RB1, MYC* or *CCND1* were found to be altered early in the evolutionary history of the tumour progression indicating that are required for tumour initiation. Additionally, the clones seeding the metastasis continue to accumulate alterations as seen by aberrations private to the metastasis and include deletions of *PTEN* and *NCOR1* or gains of *AKT2* and *CCND1*. Such aberrations have been previously described being associated with endocrine resistance like in the deletion in *PTEN*[26] or with migration and invasion like the amplification of *AKT2*.[27]

In parallel to the clonal relationship of primary tumour and metastases, the accumulation of additional mutations in the metastasis has been previously described in numerous studies.[28–30] By this definition, one would expect that in the corresponding phylogenetic tree, the metastatic branch length would be longer than the primary branch, owing to the additional accumulated aberrations. Here, we showed that this is not always the case, where approximately one third of the patients showed shorter metastasis branch length than their corresponding primary branch length, with the majority of those patients having liver and bone metastases, being de novo metastatic and having higher stromal TILs concentration in the primary tumour. In addition, copy number aberrations enriched in the "negative" class patients were also detected, including loss of *AKT1* and *NF1* and gain of *FGFR2*. The sum of the above results could point in the direction of a form of tumour dormancy. Indeed, disseminated tumour cells can migrate to the bone marrow and stay for years in a state of tumour dormancy, manifesting either as cellular or tumour mass dormancy.[31] Entering and exiting dormancy can be regulated and affected by various mechanisms that can include proliferation rate, immune surveillance, an altered transcriptional program and/or resistance to endocrine treatment.[32] Although our results cannot provide direct evidence of dormancy, they point in the direction of a decelerated tumour progression for some patients. Finally,

confirming previous reports,[33] the signal of late dissemination as portrayed through the accumulation of most of the aberrations in the trunk of the phylogenetic tree, indicates that CNAs are acquired early but continue to accumulate throughout the molecular time.

Our study has some limitations. For most of the patients, only a single metastatic lesion was biopsied. This could impact any analysis of intra-metastatic heterogeneity through the absence of information of genomic alterations in other regions of the biopsied metastasis or from other metastatic sites, including clones with acquired treatment resistance.[24,34] Furthermore, the phylogenetic trees were inferred from fully clonal CNA without considering subclonal events. This could lead to an alternative explanation of the decelerated tumour progression observed in a subset of patients, where higher clonal heterogeneity in the primary tumour could be misinterpreted as more CNA events and subsequently as a negative genetic distance. Moreover, all analyses are based on molecular time, which is assumed to be related to real time under the molecular clock hypothesis. However, it should not be overlooked that some processes may lead to the accumulation of a large number of CNAs in only a few cell cycles.[35] In this case, there could be a decoupling between the molecular and real time, that may also explain the negative genetic distances observed. Finally, we acknowledge the small number of patients in this cohort since paired primary tumour and metastatic samples are essential for such a study. As an effect of the small cohort size, the lack of statistical significance in some of the analyses performed could be attributed to the lack of statistical power. Such limitations emphasise the importance of engaging patients in research through their contribution in donating samples and genomic data that could lead to establishing large cohorts with data from multiple regions of the primary tumours and the metastatic sites.

Overall, our results validate the manifestation of intratumour heterogeneity and point towards different mechanisms of tumour evolution. As it has already been described and discussed, the evidence of late dissemination in different types of cancer can have important clinical implications. Indeed, the evidence of late dissemination further highlights the importance of early detection through cancer screening, which can be challenging in lobular breast cancer,[36] as well as the role of treatment and surgical excision of the primary tumour with the goal of preventing metastasis. Additionally, since primary and metastatic samples share a number of mutations, the primary tumour could serve as a proxy for metastasis and could guide potential treatment strategy.[33,37] At the same time, evidence of decelerated tumour progression or tumour dormancy, as demonstrated by our work, further adds to the existing discussion of how dormant disseminated tumour cells should be treated in the clinic.[38–40] Confirming these results in larger cohorts, where potential associations with clinical features and outcomes could be better explored, is necessary and could potentially influence the clinical management of ILC.

## Contributors

C.S. and D.N.B. conceived and designed the study. D.F. and D.V. performed the analyses. C.M., F.B., O.M., M. C., G.B., F.C., E.B., A.V.S., G.P. provided the clinical specimens. G.E., R.S., D.L., C.G. and G.P. performed the histopathological assessment of the samples. S.M., G.R. processed the samples. D.F., D.V., M.R., D.N.B., B. B., M.M., D.L., C.D., F.R. and C.S. interpreted the results. D.F., D.V., M.R., F.R. and C.S. wrote the manuscript. C.S. and F.R. supervised the study. D.F., DV. and C.S. have verified the underlying data. All the authors read and approved the final manuscript.

## Data sharing statement

Genome data has been deposited at the European Genome-phenome Archive (EGA) which is hosted at the EBI and CRG, under accession number EGAD00001006393, as previously described.[9] All processed data required to reproduce this analysis are available in the Supplementary Tables.

## Declaration of interests

C.S.: advisory board (receipt of honoraria or consultations fees): Astellas, Cepheid, Vertex, Seattle genetics, Puma, Amgen. Participation in company sponsored speaker's bureau: Eisai, Prime Oncology, Teva, Foundation Medicine. Other support (travel, accommodation expenses): Roche, Genentech, Pfizer (outside the submitted work). C.M. has received personal consultancy fees from Bayer, Roche, AstraZeneca, Daiichi Sankyo outside the scope of the submitted work; Grants from Italian Association for Cancer Research (AIRC); honoraria for lectures, presentations from Novartis. R.S. reports funding for research not related to the current manuscript from Breast Cancer Research Foundation; honoraria for lectures unrelated to the current manuscript from Bristol Myers Squibb; congress registration and travel not related to the current manuscript from Merck and Bristol Myers Squibb; participation on advisory boards unrelated to the current manuscript from Roche and Bristol Myers Squibb. R.S. has no conflicts of interests related to this project. G.E. reports consulting fees as general medical affairs consultancy from Thermofisher; role as secretary on the board of the scientific working group of the Belgian Society of Pathology and board member of the Belgian Society of Pathology not reimbursed. All other authors have no conflicts of interest.

## Acknowledgements

The authors would like to extend their gratitude to the patients and their families for donating clinical samples and for supporting cancer research.

This work was supported by Les Amis de l'Institut Bordet, MEDIC and the Breast Cancer Research Foundation (BCRF). D.F., M.R. and C.S. are funded by the Belgian Fonds National de la Recherche Scientifique (F.R.S-FNRS). D.V. is partly supported by a grant of the Région Wallonne.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104169.

### References


1 Guarneri V, Conte P. Metastatic breast cancer: therapeutic options according to molecular subtypes and prior adjuvant therapy. *Oncologist*. 2009;14:645–656. https://doi.org/10.1634/theoncologist.2009-0078.

2 Venet D, Fimereli D, Rothé F, et al. Phylogenetic reconstruction of breast cancer reveals two routes of metastatic dissemination associated with distinct clinical outcome. *E Bio Medicine*. 2020;56:102793. https://doi.org/10.1016/j.ebiom.2020.102793.

3 Ullah I, Karthik G-M, Alkodsi A, et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *J Clin Invest*. 2018;128:1355–1370. https://doi.org/10.1172/JCI96149.

4 Colleoni M, Sun Z, Price KN, et al. Annual hazard rates of recurrence for breast cancer during 24 years of follow-up: results from the international breast cancer study group trials I to V. *J Clin Oncol*. 2016;34:927–935. https://doi.org/10.1200/JCO.2015.62.3504.

5 Pedersen RN, Esen BÖ, Mellemkjlr L, et al. The incidence of breast cancer recurrence 10-32 years after primary diagnosis. *JNCI J Natl Cancer Inst*. 2021;114:391–399. https://doi.org/10.1093/jnci/djab202.

6 Reed AEM, Kutasovic JR, Lakhani SR, Simpson PT. Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. *Breast Cancer Res*. 2015;17:12. https://doi.org/10.1186/s13058-015-0519-x.

7 Desmedt C, Zoppoli G, Gundem G, et al. Genomic characterization of primary invasive lobular breast cancer. *J Clin Oncol*. 2016;34:1872–1881. https://doi.org/10.1200/JCO.2015.64.0334.

8 Ciriello G, Gatza ML, Beck AH, et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell*. 2015;163:506–519. https://doi.org/10.1016/j.cell.2015.09.033.

9 Richard F, Majjaj S, Venet D, et al. Characterization of stromal tumor-infiltrating lymphocytes and genomic alterations in metastatic lobular breast cancer. *Clin Cancer Res*. 2020;26:6254–6265. https://doi.org/10.1158/1078-0432.CCR-20-2268.

10 Pareja F, Ferrando L, Lee SSK, et al. The genomic landscape of metastatic histologic special types of invasive breast cancer. *Npj Breast Cancer*. 2020;6:1–10. https://doi.org/10.1038/s41523-020-00195-4.

11 Sokol ES, Feng YX, Jin DX, et al. Loss of function of NF1 is a mechanism of acquired resistance to endocrine therapy in lobular breast cancer. *Ann Oncol*. 2019;30:115–123. https://doi.org/10.1093/annonc/mdy497.

12 Desmedt C, Pingitore J, Rothé F, et al. ESR1 mutations in metastatic lobular breast cancer patients. *NPJ Breast Cancer*. 2019;5:9. https://doi.org/10.1038/s41523-019-0104-z.

13 Salgado R, Denkert C, Demaria S, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol*. 2015;26:259–271. https://doi.org/10.1093/annonc/mdu450.

14 Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26:589–595. https://doi.org/10.1093/bioinformatics/btp698.

15 Picard Tools - By Broad Institute n.d. http://broadinstitute.github.io/picard/. Accessed 17 April 2021.

16 Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*. 2016;12:e1004873. https://doi.org/10.1371/journal.pcbi.1004873.

17 Nilsen G, Liestøl K, Van Loo P, et al. Copynumber: efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics*. 2012;13:591. https://doi.org/10.1186/1471-2164-13-591.

18 Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30:413–421. https://doi.org/10.1038/nbt.2203.

19 El-Kebir M, Raphael BJ, Shamir R, et al. Complexity and algorithms for copy-number evolution problems. *Algorithms Mol Biol*. 2017;12:13. https://doi.org/10.1186/s13015-017-0103-2.

20 Nik-Zainal S, Davies H, Staaf J, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534:47–54. https://doi.org/10.1038/nature17676.

21 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313.

22 Hu Z, Ding J, Ma Z, Sun R, et al. Quantitative evidence for early metastatic seeding in colorectal cancer. *Nat Genet*. 2019;51:1113–1122. https://doi.org/10.1038/s41588-019-0423-x.

23 Reiter JG, Hung W-T, Lee I-H, et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat Genet*. 2020;52:692–700. https://doi.org/10.1038/s41588-020-0633-2.

24 Brown D, Smeets D, Székely B, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun*. 2017;8:14944. https://doi.org/10.1038/ncomms14944.

25 Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*. 2012;366:883–892. https://doi.org/10.1056/NEJMoa1113205.

26 Fu X, Creighton CJ, Biswal NC, et al. Overcoming endocrine resistance due to reduced PTEN levels in estrogen receptor-positive breast cancer by co-targeting mammalian target of rapamycin, protein kinase B, or mitogen-activated protein kinase kinase. *Breast Cancer Res*. 2014;16:430. https://doi.org/10.1186/s13058-014-0430-x.

27 Hinz N, Jücker M. Distinct functions of AKT isoforms in breast cancer: a comprehensive review. *Cell Commun Signal*. 2019;17:154. https://doi.org/10.1186/s12964-019-0450-3.

28 Brastianos PK, Carter SL, Santagata S, et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov*. 2015;5:1164–1177. https://doi.org/10.1158/2159-8290.CD-15-0369.

29 Yates LR, Gerstung M, Knappskog S, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med*. 2015;21:751–759. https://doi.org/10.1038/nm.3886.

30 Savas P, Teo ZL, Lefevre C, et al. The subclonal architecture of metastatic breast cancer: results from a prospective community-based rapid autopsy program "CASCADE. *PLoS Med*. 2016;13:e1002204. https://doi.org/10.1371/journal.pmed.1002204.

31 Aguirre-Ghiso JA. Models, mechanisms and clinical evidence for cancer dormancy. *Nat Rev Cancer*. 2007;7:834–846. https://doi.org/10.1038/nrc2256.

32 Riggio AI, Varley KE, Welm AL. The lingering mysteries of metastatic recurrence in breast cancer. *Br J Cancer*. 2021;124:13–26. https://doi.org/10.1038/s41416-020-01161-4.

33 Yates LR, Knappskog S, Wedge D, et al. Genomic evolution of breast cancer metastasis and relapse. *Cancer Cell*. 2017;32:169–184.e7. https://doi.org/10.1016/j.ccell.2017.07.005.

34 Razavi P, Chang MT, Xu G, et al. The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer Cell*. 2018;34:427–438.e6. https://doi.org/10.1016/j.ccell.2018.08.008.

35 Umbreit NT, Zhang C-Z, Lynch LD, et al. Mechanisms generating cancer genome complexity from a single cell division error. *Science*. 2020;368:eaba0712. https://doi.org/10.1126/science.aba0712.

36 Christgen M, Steinemann D, Kühnle E, et al. Lobular breast cancer: clinical, molecular and morphological characteristics. *Pathol - Res Pract*. 2016;212:583–597. https://doi.org/10.1016/j.prp.2016.05.002.

37 Leung ML, Davis A, Gao R, et al. Single-cell DNA sequencing reveals a late-dissemination model in metastatic colorectal cancer. *Genome Res*. 2017;27:1287–1299. https://doi.org/10.1101/gr.209973.116.

38 Ghajar CM. Metastasis prevention by targeting the dormant niche. *Nat Rev Cancer*. 2015;15:238–247. https://doi.org/10.1038/nrc3910.

39 Aguirre-Ghiso JA, Bragado P, Sosa MS. Metastasis awakening: targeting dormant cancer. *Nat Med*. 2013;19:276–277. https://doi.org/10.1038/nm.3120.

40 Shimizu H, Takeishi S, Nakatsumi H, Nakayama KI. Prevention of cancer dormancy by Fbxw7 ablation eradicates disseminated tumor cells. *JCI Insight*. 2019;4(4):e125138. https://doi.org/10.1172/jci.insight.125138.