



Research paper

Phylogenetic reconstruction of breast cancer reveals two routes of metastatic dissemination associated with distinct clinical outcome

David Venet^{a,1}, Danai Fimereli^{a,1}, Françoise Rothé^{a,1}, Bram Boeckx^{b,c}, Marion Maetens^{a,d}, Samira Majjaj^a, Ghizlane Rouas^a, Maria Capra^e, Giuseppina Bonizzi^e, Federica Contaldo^e, Christine Galant^f, Martine Piccart^g, Giancarlo Pruneri^{e,h,i}, Denis Larsimont^j, Diether Lambrechts^{b,c}, Christine Desmedt^{a,d,1}, Christos Sotiriou^{a,1,*}

^aJ.C. Heuson Breast Cancer Translational Research Laboratory, Institut Jules Bordet, Université Libre de Bruxelles, Bld de Waterloo 121, 1000, Brussels, Belgium

^bLaboratory of Translational Genetics, VIB Center for Cancer Biology, Campus Gasthuisberg, O&N IV Herestraat 49 - box 912, 3000, Leuven, Belgium

^cLaboratory of Translational Genetics, Department of Human Genetics, Katholieke Universiteit Leuven, O&N IV Herestraat 49 - box 912, 3000, Leuven, Belgium

^dLaboratory for Translational Breast Cancer Research, Department of Oncology, Katholieke Universiteit Leuven, O&N I Herestraat 49 - box 818, 3000, Leuven, Belgium

^eBiobank for Translational Medicine Unit, Department of Pathology, European Institute of Oncology, Via Ripamonti 435, 20141, Milan, Italy

^fDepartment of Pathology, Cliniques Universitaires Saint Luc, Avenue Hippocrate 10, 1200, Brussels, Belgium

^gMedical Oncology Department, Institut Jules Bordet, Université Libre de Bruxelles, Bld de Waterloo 121, 1000, Brussels, Belgium

^hDivision of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori, via Venezian 1, 20133, Milan, Italy

ⁱSchool of Medicine, University of Milan, via Festa del Perdono 7, 20122, Milano, Milan, Italy

^jDepartment of Pathology, Institut Jules Bordet, Bld de Waterloo 121, 1000, Brussels, Belgium



ARTICLE INFO

Article History:

Received 16 December 2019

Revised 24 April 2020

Accepted 27 April 2020

Available online xxx

Keywords:

Breast cancer

Metastatic dissemination

Axillary lymph node

Distant metastasis

ABSTRACT

Background: In breast cancer (BC), axillary lymph node (ALN) involvement is one of the strongest adverse prognostic factors. However, it is unclear whether loco-regional lymph node deposits are effectively the root of secondary metastases or only an indicator of competence of the primary tumour to spread to distant organs.

Methods: Here, we investigated the evolutionary trajectories of primary tumour, ALN and distant metastasis samples from 16 estrogen-receptor (ER)-positive lymph node-positive BC patients. Low-pass whole genome sequencing was performed to infer somatic copy number aberrations and the phylogenetic profiles for all patients were obtained.

Findings: We show that lymph nodes and distant metastases shared a common origin in only 25% of the cases highlighting that the predominant route of metastatic dissemination is the direct seeding of tumour cells from the primary tumour to distant organs, independently of lymph node metastasis. Noticeably, patients sharing a common origin significantly have worse prognosis.

Interpretation: Our results shed light on the routes on which tumour cells metastasize and their role in disease progression in ER-positive BC.

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

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Local and regional lymph node metastases are associated with poor prognosis in a variety of cancers [1]. Based on the hypothesis that axillary lymph node metastases constitute a reservoir for further

dissemination to distant organs, complete surgical extirpation of axillary lymph nodes has been the standard practice for the treatment of primary breast cancers presenting involved lymph nodes for almost a century [2,3]. However, later research demonstrated that omission of initial axillary resection did not adversely affect breast cancer mortality, suggesting that dissemination occurs not only via the axillary lymphatic system but also through an alternative route [4]. In colorectal cancer, Naxerova *et al.* found that in 65% of patients, regional lymph nodes had a distinct origin from distant organ

* Corresponding author.

E-mail address: christos.sotiriou@bordet.be (C. Sotiriou).

¹ These authors contributed equally

Research in context

Evidence before this study

In breast cancer, cancer cells from the primary tumour site can disseminate and reach distant organs through two different routes. These routes involve either the blood circulation or the lymphatic vessels. The lymph node involvement is an integral component of the tumour staging and plays an important role in therapeutic strategies. However, it still remains debated whether lymph node metastases can seed distant metastases or whether they merely represent a surrogate marker for tumour aggressiveness.

Added value of the study

We found that the direct seeding of tumour cells from the primary tumour to distant organs is the main route of metastatic dissemination in ER-positive breast cancer. In parallel, the dissemination route involving the lymph nodes was found only in a small number of patients and most importantly was associated with a worse prognosis.

Implications of the available evidence

Our work underlines the routes on which tumour cells metastasize in breast cancer. The differences observed between these distinct dissemination routes associated with disease progression could impact the clinical management of early lymph node-positive breast cancer patients.

involvement at the time of primary tumour diagnosis and a confirmed distant organ metastasis of breast cancer origin as per medical records. Eligible patients were further narrowed down to include only those for whom at least one primary, one axillary lymph node and one distant metastatic as well as a histologically normal tissue sample as germline reference were available as formalin-fixed paraffin-embedded (FFPE) tissue blocks. All patients were diagnosed between 1994 and 2012 and treated with surgery, followed by medical treatment as per local guidelines at the time of presentation. The extended clinico-pathological characteristics of the patients is provided in Table S2.

2.2. Histopathological characterization and DNA extraction

Histopathological evaluation of the FFPE tissue samples was carried out independently by two pathologists and involved histological subtyping, assessment of histological grade, and determination of the percentage of tumour epithelial cells. Whenever the estimated tumour cellularity was below 50%, macrodissection was performed in order to enrich the samples in the tumour epithelial cells. Immunohistochemical evaluations of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) and Ki67 were retrieved from local routine assessments. DNA extractions from FFPE samples were performed using the QIAamp DNA FFPE Tissue Kit (Qiagen). The concentration of double stranded DNA was measured using the Qubit fluorometer (Invitrogen). Only patients for whom > 100 ng of double stranded DNA was available for at least one sample each of a histologically normal, primary tumour, axillary and distant metastatic sample, were considered for downstream whole genome sequencing.

2.3. 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.

2.4. Bioinformatics analyses

The raw sequence reads were aligned to the human genome reference hg19/GRCh37 using the BWA aligner [9] and duplicate reads were marked using Picard [10] resulting in approximately 9M aligned reads per sample. 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 15 kbp windows and corrected for library size and GC content using CNVkit [11]. 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.* [12] 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 [13] in order to infer the cancer cell fraction (CCF), genomic mass, and segment wise copy numbers using the following equation:

metastases, raising the possibility that the latter were seeded independently [5]. In the remaining 35% of cases, distant organ metastases were monophyletic with at least one local lymph node deposit, suggesting that cancer cells could also disseminate from nodes to distant organ sites. This distinction is also of relevance to the treatment of breast cancers where ALN involvement is an integral component of the TNM staging system [6]. In support of the multiple origins of ALN and distant metastases, two recent studies using mice models of breast cancer showed that tumour cells can disseminate through the bloodstream to form distant metastases by extravasating to local blood rather than through lymphatic vessels as previously thought [7,8].

Here, using matched primary tumour, axillary lymph node and distant metastatic samples from breast cancer patients, we explored to which extent ALN are transit points for further dissemination events to distant organs and assessed their role in disease progression. We identified two different routes of metastatic dissemination, either through the lymph node or directly from the primary tumour to the distant metastasis. Our findings highlight that the predominant route of metastatic dissemination operating in ER-positive breast cancer is the direct seeding of tumour cells from the primary tumour to distant organs, independently of lymph node metastasis, shedding light on the routes on which tumour cells metastasize and their role in disease progression in ER-positive BC patients.

2. Materials and Methods

2.1. Patients' selection and samples' collection

We retrospectively identified a cohort of breast cancer patients from the tissue banks of the Institut Jules Bordet (Brussels, Belgium), the Cliniques Universitaires Saint Luc (Brussels, Belgium) and the European Institute of Oncology (Milan, Italy). Inclusion criteria were patients with primary breast tumour, axillary lymph node

$$y = \frac{1}{\alpha} \left(2^{\frac{x}{c}} (\psi\alpha + 2(1 - \alpha)) - 2(1 - \alpha) \right)$$

where x is the segmented \log_2 ratio of a particular genomic locus, α is the CCF, ψ is the genomic mass, and c is the compression ratio which is set to 1 in the case of whole genome sequence based \log_2 ratios. 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 y of copy numbers and rounded them to the nearest integer value. These values were used as input to CNT-ILP [14], run with 25G maximum memory and 40 h maximum time on 4 cores. All other parameter settings were kept at default 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. Further visualization was done in R.

2.5. Statistical analysis

Statistical support for the phylogenetic trees were 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. Each of the reconstructed bootstrap tree was then classified either as belonging to the distinct or the common origin, similar to the original classification rules. The bootstrap percentage values then correspond to the number of phylogenetic trees which are correctly classified out of the 50 bootstrap replicates. We defined gains or amplifications as segments with CN above 1.25 or 2 times the genomic mass of the sample, and losses as segments with CN below 1.5. Statistical hypothesis tests and associations with patient clinico-pathological characteristics and types of aberration involving the comparison of two or more groups were carried out using the Mann-Whitney U-test or Fisher's exact test as appropriate. Survival data are shown using the Kaplan-Meier estimator, while the Cox regression analysis was used to calculate the associated p-value, correcting for the nodal status (N1 vs N2+) and grade (G1/2 vs G3). Unless otherwise specified, p-values were unadjusted.

2.6. Ethics statement

This study received approval from the respective institutional ethics committees.

3. Results

3.1. Patterns of local and distant metastatic dissemination

A total of 235 samples from 30 lymph node-positive breast cancer patients were profiled using low coverage whole genome sequencing to infer somatic copy number aberrations (CNA). Given that axillary lymph node metastases are more frequent in ER-positive tumours [15,16,17,18,19,20], only patients with ER-positive primary breast cancer were included in the study. After rigorous filtering and processing of the sequencing data, the integer estimates of CNA were used to reconstruct the phylogenetic trees representing metastatic progression from 16 patients with at least one primary tumour, one positive lymph node and one distant metastatic sample (see Methods, Table 1, Fig. S1 and S4 and Table S1). In contrast to existing reports with only a few ALN metastasis examined [21,22,23,24], the presence of multiple ALN and distant metastases in our cohort allowed us to explore in depth the possible dissemination routes. The evolutionary relationship between lymphatic and distant metastases categorized patients based on two separate trajectories: those whose

Table 1
Clinical and pathological characteristics of patients (N = 30)

Characteristic	Number of patients (%)
<i>Age at diagnosis (years)</i>	
<50	12 (40)
≥50	18 (60)
<i>Size (cm)</i>	
<2	4 (13)
≥2	24 (80)
Unknown	2 (7)
<i>PgR status[‡]</i>	
Positive (+ve)	26 (87)
Negative (-ve)	4 (13)
<i>HER2 status[‡]</i>	
Positive (+ve)	2 (7)
Negative (-ve)	27 (90)
Unknown	1 (3)
<i>Histological grade</i>	
1	3 (10)
2	17 (60)
3	10 (30)
<i>Ki67 (%)</i>	
<20	19 (63)
≥20	11 (37)

Unless otherwise stated, all clinico-pathological features refer to the primary breast tumour † PgR; Progesterone receptor, HER2; Human epidermal growth factor receptor 2

metastases have a distinct origin from lymph nodes and those whose metastases have a common origin with lymph nodes (Fig. 1a, d). More specifically, the first category implies dissemination from the primary tumour to the axilla and a distinct seeding event from the primary tumour to a distant organ. In this case, the primary tumour is genetically closer either to the axillary lymph node metastasis or to the distant lesion. In the second category, the axillary lymph node deposit has a more recent common ancestry with the distant metastasis implying that the latter may have been seeded through the axillary lymphatic system. This classification is consistent with previous reports, which revealed a linear progression in most patients, while parallel seeding events from the primary tumour were also observed in some patients [25,21,26,27,28,29]. The phylogenetic reconstruction revealed that 12 (75%) patients fell into the distinct origin category while the second common origin category, where ALN are monophyletic with the distant lesions, was found in 4 (25%) patients, namely PT02, PT06, PT13 and PT15 (Fig. S2) highlighting that both routes of dissemination are possible in ER-positive breast cancer patients. However, lymph nodes and distant metastases shared a common origin in only 25% of the cases, implying that the lymph node is an intermediate that seeds to the distant metastasis while for the majority of the patients, the distant metastasis seeded directly from the primary tumour indicating a distinct origin. Selected patients from both origins are illustrated in Figure 1. Patient PT22 is an example of the distinct origin dissemination route where multiple subclones seeded genetically distinct lymph node and metastasis. Clones from the primary tumour sample 1 (P1) were genetically closer to the lymph node metastasis, whereas primary tumour sample 3 (P3) was closer to the liver distant metastasis. In the 4 cases with common origin dissemination route, the lymph nodes and the distant metastases were genetically closer and seeded from the primary tumour in a sequential manner. Similar findings were observed in colorectal cancer [5]. Further noteworthy examples include patients PT21 and PT24, both falling into the distinct dissemination origin category (Fig S2). In patient PT24, clones from the primary tumour were phylogenetically closer to the lymph node but rather distant from the bone metastasis clone, indicating a greater heterogeneity and a possible parallel evolution model. The case of patient PT21 highlights a case of different clone evolution, one that led to the lymph node metastases PLN1 and PLN2 (clone P1) and one that led to the distant bone metastasis (M1) and the lymph node metastasis

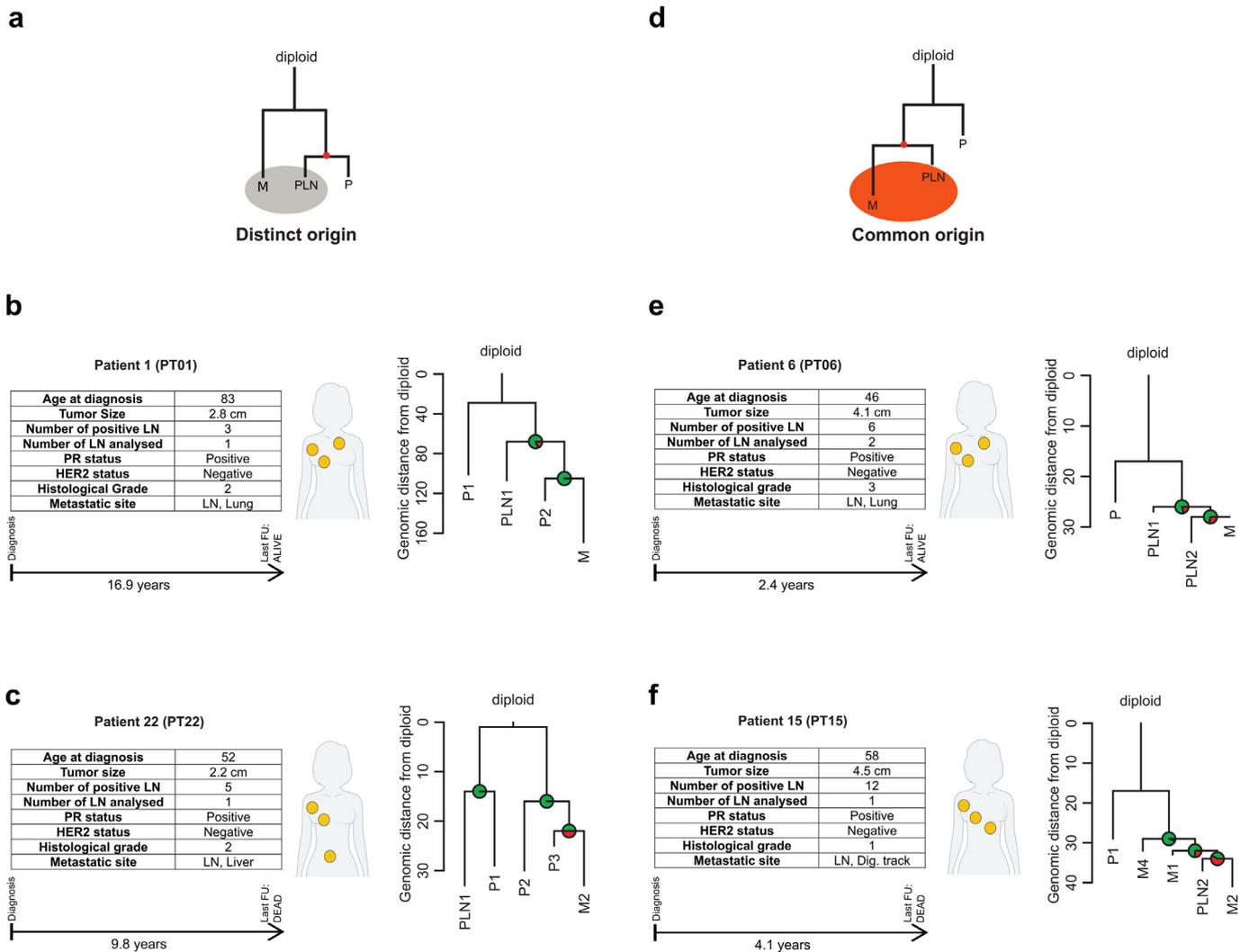


Fig. 1. Phylogenetic trees of patients with distinct and common origins of lymphatic and distant metastases. (a) and (d) model tree topologies of the distinct and the common origin classification. In the distinct origin classification model, the primary tumour (P) seeds independently the lymph node (PLN) and the distant metastasis (M). In the common origin classification model, the distant metastasis is seeded from the lymph node. (b)–(c) phylogenetic trees of representative patients PT01 and PT22 classified in the distinct origin. (e)–(f) phylogenetic trees of representative patients PT06 and PT15 classified in the common origin. In (b), (c), (e) and (f) the pie charts in each branch indicate how often this branch appears in the bootstraps. A green filled pie indicates that this branch appeared 100% of the times in the bootstraps. Clinical information boxes show the patient's age at diagnosis, tumour size, number of positive lymph nodes (LN), number of lymph nodes analysed, progesterone receptor (PR) and HER2 status, histological grade and metastatic sites.

PLN3 (clone P2). Interestingly, we also identified two patients (PT19 and PT23) with distant bone metastasis, where the metastatic samples have accumulated less aberrations than their primary tumour, possibly reflecting a period of quiescence or dormancy.

3.2. The number of samples affects the reconstruction of the evolutionary trajectories

To ascribe a statistical confidence to our classification, we performed two complementary analyses by permuting our data at different levels, firstly through bootstrapping of the samples and secondly through bootstrapping of the segmented copy number data. In the first analysis, patients were divided into three groups: a group consisting of patients with multiple primary samples ($n = 10$ patients), a second group consisting of patients with multiple lymph node samples ($n = 8$ patients) and a third group consisting of patients with multiple metastatic samples ($n = 2$ patients). Then, for each group we kept only a single sample (instead of all analysed samples), recalculated the phylogenetic trees and reclassified patients according to the two dissemination routes (Table S4). Figure 2a-c shows the change in the origin classification when we only kept a single primary, lymph node or metastatic sample in each group respectively. During the reconstruction of the phylogenetic

trees in the first group, we observed that 40% of the patients were classified, at least once, in a different dissemination route than that of the original classification. This change in the classification can be seen through the example of patient PT01, classified originally in the distinct origin category (Fig. 2e), where the removal of the primary tumour sample 2 (P2) clusters the metastasis sample and the lymph nodes together, thus changing the origin classification (Fig. 2f). When recalculating the trees in the second group, we observed a change in the classification in a smaller proportion of patients (25%), illustrated with the example of patient PT25, originally classified in the distinct origin group (Fig. 2g) and who changed classification when removing the multiple lymph node samples (Fig. 2h). Finally, we did not observe a change of origin for the patients in the third group with multiple metastatic samples. Since the current analysis did not consider subclonal events, one cannot exclude the possibility that rare metastasis-competent clones from the primary tumour seeded the axillary lymph nodes and distant lesions in parallel, therefore changing the classification of common origin. In the second analysis, the underlying CNA data were resampled and the phylogenetic trees were inferred from the bootstrapped data. Overall, for 14 out of 16 patients, trees with identical topology as the original phylogeny were obtained after bootstrapping the data as witnessed by a $> 80\%$ bootstrap confidence (Fig. 2d), emphasizing the robustness of our

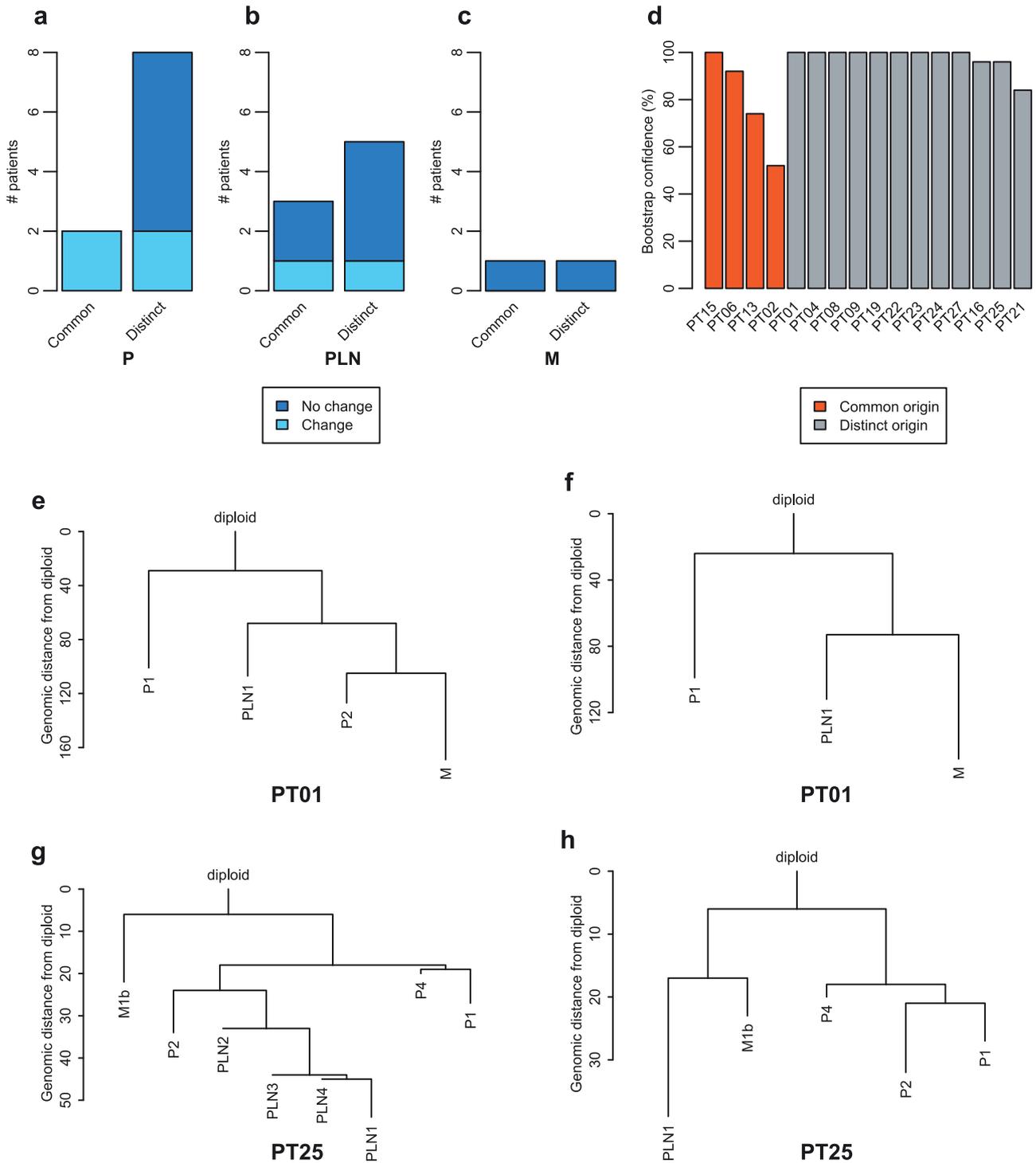


Fig. 2. Bootstrapping results of the phylogenetic tree reconstruction. Number of patients with the same or different tree topology after removal of (a) multiple primary samples (Number of patients N = 10), (b) multiple lymph node samples (Number of patients N = 8) or (c) multiple metastatic samples (Number of patients N = 2), (d) percentage of phylogenetic trees recovered in 50 bootstrap replicates with identical topology to the original phylogeny, (e) - (f) original phylogenetic tree of patient PT01 containing all the samples and new phylogenetic tree of the same patient after removal of multiple primary samples, (g-h) original phylogenetic tree of patient PT25 containing all the samples and new phylogenetic tree of the same patient after removal of multiple lymph node samples.

methodology (Table S5). It should not be overlooked that the number of samples could influence the reconstruction of any phylogenetic tree. As shown in our analysis, this known caveat is also observed during the reconstruction of the evolutionary trajectories of cancer patients, where the number of primary and ALN samples profiled per patient can influence the origin classification, underlying thus the importance of multi-region sampling in such studies.

3.3. Common origin dissemination route is associated with worse outcome

We then examined the clinico-pathological variables associated with the origin classification and showed that the common origin dissemination route was significantly associated with worse overall survival (Fig. S3), with the caveat that sample sizes are too small to

obtain robust results. We also observed a non-significant trend between this route and larger tumour size as well as low histological grade tumours (Fig. 3). These results come in contrast to Naxerova *et al.* where no difference of any clinico-pathological variable between the origins was reported [5]. Apart from the aforementioned associations, we did not observe any significant difference in age, time to recurrence, Ki67 level, nodal status, number of positive lymph nodes or number of positive lymph nodes for which we were able to estimate a CNA profile between the two origin categories (Fig. 3). The association of a higher tumour size with the common origin could reflect genuine biology, indicating an association with patient survival. Of course, as large tumours are genetically more heterogeneous, it is possible that only part of the specimen from surgical resection would have been assayed in our study therefore leading to a different classification of origin when in fact, rare metastasis competent clones seeded in parallel both the lymph nodes and distant metastasis [30]. An unexpected finding in our study was however that patients with lymph node to metastasis dissemination events in our cohort were mostly low grade (1 and 2). Interestingly, these common origin patients were found to have significantly worse prognosis than the remaining patients, despite being low grade. The above results could imply that lymph nodes are selectors of more aggressive clones that could later disseminate to distant sites and thus lead to worse survival [31].

3.4. MYC amplification differentiates the metastatic dissemination routes

As chromosomal instability (CIN) is associated with tumour metastasis and is frequently found in metastatic clones [32], we questioned whether it could also affect the selection of the dissemination route to distant metastasis. We did not find any significant association between the fraction of genome altered and the route of dissemination (Fig. 4a). We further explored whether differences in CNA deletions, gains and amplifications were associated with the origin classification (Fig. 4a), focusing on the 31 known breast cancer copy number driver genes [33]. We have shown that *CCND1* (62%), *MYC* (50%) and *FGFR1* (44%) genes were the most frequently gained/amplified genes whereas *CDH1* (81%), *TP53* (62%), *MAP2K4* (56%) and *NCOR1* (56%) genes were the most frequently deleted genes across the entire cohort (Fig. 4c). When comparing the two origin classifications, we observed that *MYC* amplification was the only one more prevalent in the distinct origin patients as compared to the common origin patients ($p = 0.0011$) (Fig. 4b).

4. Discussion

The mechanisms involved in the progression of cancers have been described in details [34,35,36]. However, it remains debated whether tumour cells from primary tumours reach distant organs directly

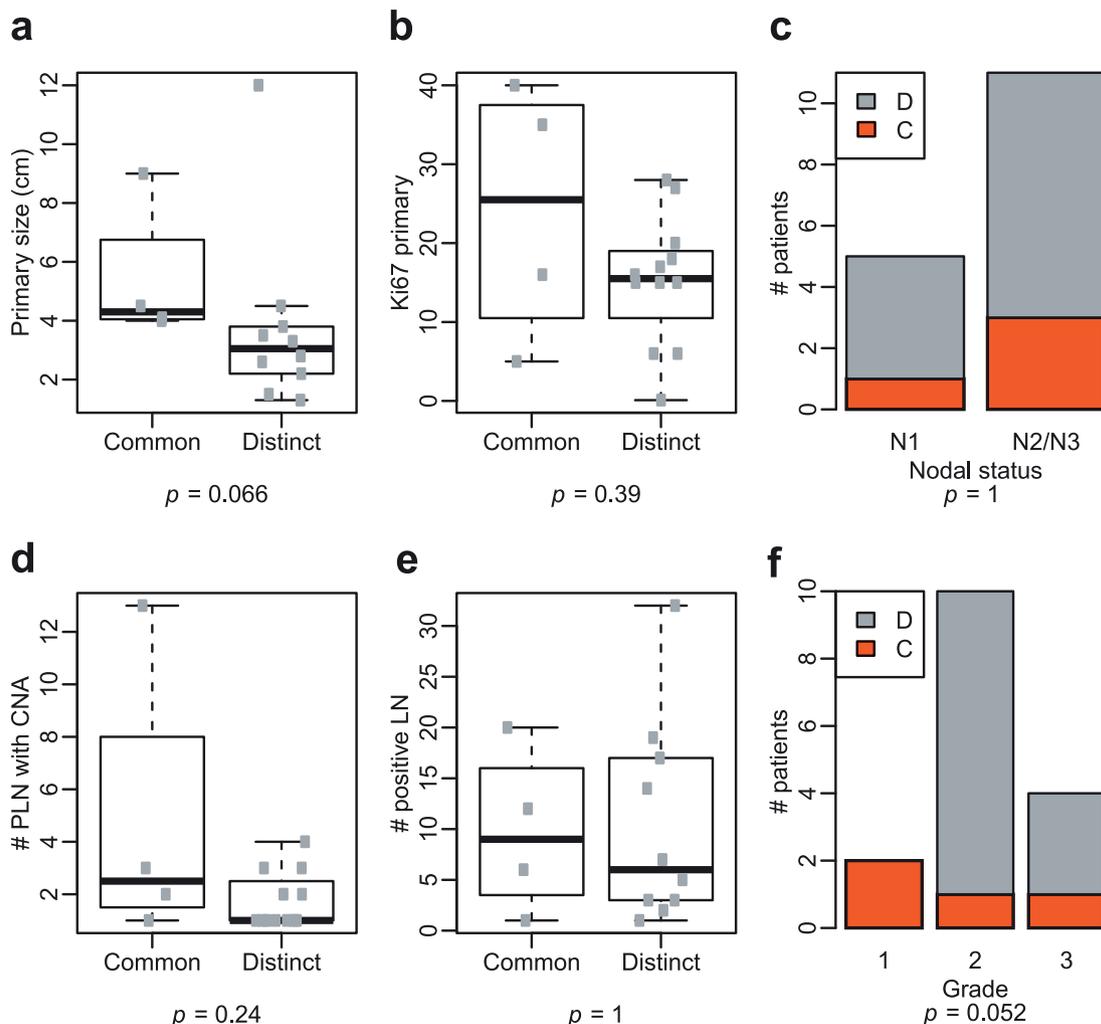


Fig. 3. Associations between distinct and common origins and clinico-pathological variables. Association of tree topologies with (a) primary tumour size, (b) Ki67 status, (c) nodal status, (d) number of positive lymph nodes (PLN) for which we were able to estimate the copy number aberrations (CNA), (e) number of positive lymph nodes (LN), (f) grade. In (a), (b), (d) and (e) the p-value was obtained using a Mann-Whitney U-test. In (c) and (f) the p-value was obtained using Fisher's exact test.

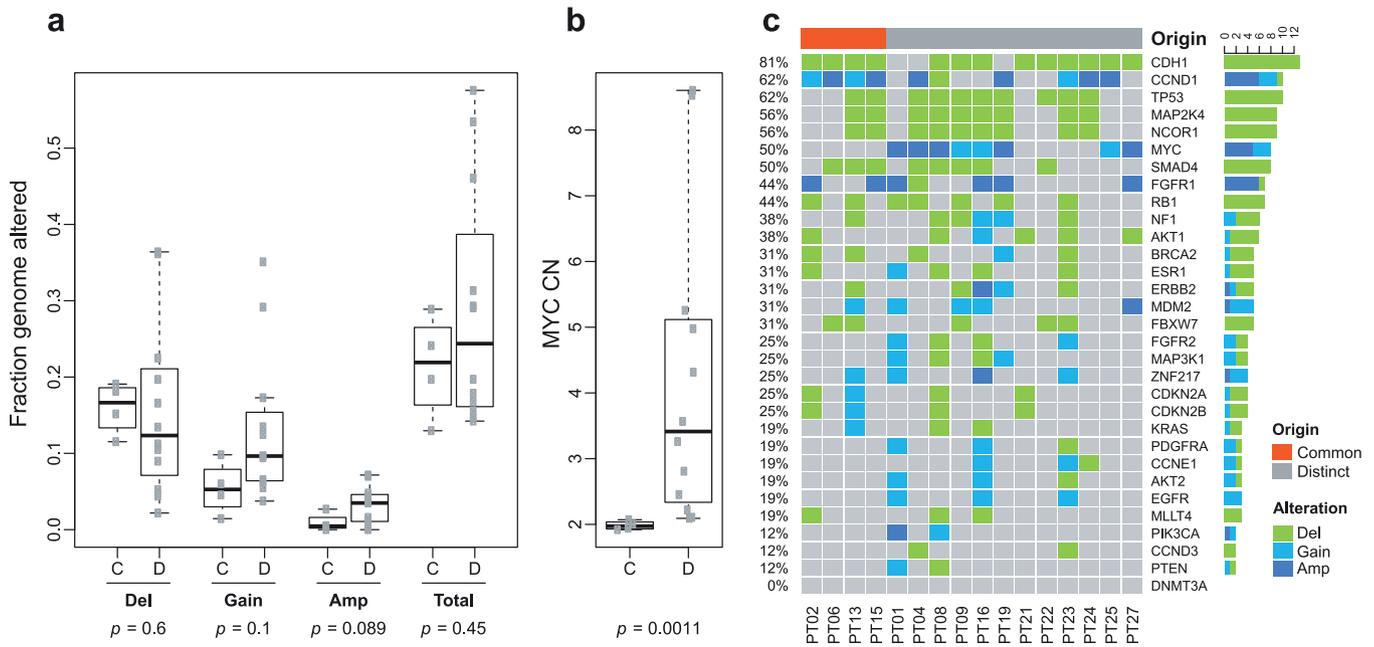


Fig. 4. Genomic instability within the different dissemination routes. (a) Fraction of genome altered by deletions, gains, amplifications, and total alterations in function of dissemination routes, (b) number of MYC copies in the common and distinct origin patients (c) heatmap showing the copy number alterations in 31 breast cancer copy number driver genes. The percentage of samples altered is shown on the left side and the frequencies of deletions, gains and amplifications are shown on the right side of the heatmap. The p-values were obtained using a Mann-Whitney U-test.

through venous capillaries or first transit through the lymph nodes. A recent study in colorectal cancer showed that in the majority of the patients distant metastases seed directly from the primary tumour [5]. In breast cancer, only a handful of studies have reported so far to the best of our knowledge the phylogenetic relationship between primary tumours, ALN and distant metastases. Three studies reported each single cases with each time only one ALN metastasis sample [22,23,24]. In all cases they found the ALN metastasis to be phylogenetically closer to the primary tumour than the distant metastasis. Additionally, the recent results of Ullal *et al.* [21], which are based on 8 patients, are consistent with the above results on the three isolated patients and therefore suggest that ALN metastases are rather an indicator of the aggressiveness of the disease rather than an intermediate step before distant dissemination. However, these studies did not account for the molecular heterogeneity of breast cancer since they included all BC molecular subtypes. Furthermore, despite multiple ALN being involved in the majority of the patients investigated in these studies, only a single ALN metastasis has been sequenced for all but two patients for which two ALN metastases were sequenced. To overcome such limitations, in this study we reconstructed the evolutionary trajectories of 16 ER-positive breast cancer patients with multiple primary, ALN and distant metastasis samples. In contrast to the above studies, we provided evidence that lymph nodes are, alongside their strong prognostic factor, an intermediate step in the dissemination cascade to distant metastasis, although it accounts for a small percentage of the cases. Interestingly, we showed that the common origin patients have worse prognosis than the distinct origin patients, despite being low grade, indicating that lymph nodes are potential selectors of more aggressive clones albeit the small sample size could impact the test statistics and this finding would need to be confirmed in larger studies. Additionally, we found that MYC amplification, was significantly more frequent in the cases where the distant metastasis seeded directly from the primary tumour, thus providing further evidence of its association with distant metastasis in breast cancer [37,38].

In conclusion, we showed that the two modes of dissemination described earlier also operate in breast cancer. Unexpectedly, the majority of the distant metastases are seeded directly from the

primary tumour independently of lymph node metastasis. The gradual de-escalation of axillary surgery in order to reduce long-term morbidity associated with this procedure [39,40], encouraged acceptance of the sentinel lymph node procedure, which had gained increased traction since the early 2000s [41], to become the standard practice for evaluating axillary involvement. However, the worse survival associated with the dissemination route passing through the lymph nodes, as demonstrated by our work, highlights the intrinsic biological properties of these aggressive disseminating clones that may potentially influence the clinical management of early ER-positive lymph node-positive breast cancer patients.

4.1. Data availability

Genome data has been deposited at the European Genome-phenome Archive (EGA) which is hosted at the EBI and CRG, under accession number EGAS00001004356. The CNA data for reproducing the analysis are available in Supplemental Table S3.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

This work has received financial support from Les Amis de l'Institut Bordet, MEDIC and the Breast Cancer Research Foundation (BCRF). CS is funded by the Belgian Fonds National de la Recherche Scientifique (F.R.S-FNRS). DV is partly supported by a grant of the Région Wallonne. CD is supported by Les Amis de l'Institut Bordet. We thank David Brown for his assistance with the bioinformatics analyses and the manuscript writing which was partly described in his PhD thesis (<http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/260251>). The funders did not have any role in study design, data collection, data analysis, interpretation or writing of the report.

Author Contributions

CS and CD conceived and designed the study. MC, GB, FC, GP provided the clinical specimens. DL, CG and GP performed the histopathological assessment of the samples. SM, GR processed the samples. DV, DF performed the analyses. DV, DF, BB, MM, FR, MP, DL, CD and CS interpreted the results. DV, DF, FR, CD and CS wrote the manuscript. CS and CD supervised the study. All the authors read and approved the final manuscript.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2020.102793.

References

- Nathanson SD. Insights into the mechanisms of lymph node metastasis. *Cancer* 2003;98:413–23.
- Halsted WS. The results of operations for the cure of cancer of the breast performed at the Johns Hopkins hospital from June, 1889, to January, 1894. *Ann Surg* 1894;20:497–555.
- Patey DH, Dyson WH. The prognosis of carcinoma of the breast in relation to the type of operation performed. *Br J Cancer* 1948;2:7–13.
- Fisher B, Jeong JH, Anderson S, Bryant J, Fisher ER, Wolmark N. Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N Engl J Med* 2002;347:567–75.
- Naxerova K, Reiter JG, Brachtel E, Lennerz JK, Van De Wetering M, Rowan A, et al. Origins of lymphatic and distant metastases in human colorectal cancer. *Science* 2017;357(80):55–60.
- Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989;63:181–7.
- Brown M, Assen FP, Leithner A, Abe J, Schachner H, Asfour G, et al. Lymph node blood vessels provide exit routes for metastatic tumor cell dissemination in mice. *Science* 2018;359(80):1408–11.
- Pereira ER, Kedrin D, Seano G, Gautier O, Meijer EFJ, Jones D, et al. Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. *Science* 2018;359(80):1403–7.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
- Picard toolkit [Internet]. Broad Institute. GitHub repository. Broad Institut; 2019. Available from: <http://broadinstitute.github.io/picard/>.
- 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.
- Nilsen G, Liestøl K, Van Loo P, Moen Volla HK, Eide MB, Rueda OM, et al. Copy-number: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* 2012;13.
- 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;30:413–21.
- El-Kebir M, Raphael BJ, Shamir R, Sharan R, Zaccaria S, Zehavi M, et al. Complexity and algorithms for copy-number evolution problems. *Algorithms Mol Biol* 2017;12:13.
- Gann PH, Colilla SA, Gapstur SM, Winchester DJ, Winchester DP. Factors associated with axillary lymph node metastasis from breast carcinoma. *Cancer* 1999;86:1511–9.
- Lu X, Lu X, Wang ZC, Iglehart JD, Zhang X, Richardson AL. Predicting features of breast cancer with gene expression patterns. *Breast Cancer Res Treat* 2008;108:191–201.
- Liu N, Yang Z, Liu X, Niu Y. Lymph node status in different molecular subtype of breast cancer: Triple negative tumours are more likely lymph node negative. *Oncotarget* 2017;8:55534–43.
- Mattes MD, Bhatia JK, Metzger D, Ashamalla H, Katsoulakis E. Breast cancer subtype as a predictor of lymph node metastasis according to the SEER registry. *J Breast Cancer* 2015;18:143–8.
- Sawaki M, Idota A, Ichikawa M, Gondo N, Horio A, Kondo N, et al. Impact of intrinsic subtype on predicting axillary lymph node metastasis in breast cancer. *Oncol Lett* 2014;8:1707–12.
- Buonomo OC, Caredda E, Portarena I, Vanni G, Orlandi A, Bagni C, et al. New insights into the metastatic behavior after breast cancer surgery, according to well-established clinicopathological variables and molecular subtypes. *PLoS One* 2017;12.
- Ullah I, Karthik GM, Alkodsai A, Kjällquist U, Stålhammar G, Lövrot J, et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *J Clin Invest* 2018;128:1355–70.
- 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;32:169–84 e7.
- Brastianos PK, Carter SL, Santagata S, Cahill DP, Taylor-Weiner A, Jones RT, et al. Genomic characterization of brain metastases reveals branched evolution and potential therapeutic targets. *Cancer Discov* 2015;5:1164–77.
- Murtaza M, Dawson SJ, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* 2015;6:8760.
- Brown D, Smeets D, Székely B, Larsimont D, Szász AM, Adnet P-Y, et al. Phylogenetic analysis of metastatic progression in breast cancer using somatic mutations and copy number aberrations. *Nat Commun* 2017;8:14944.
- Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, et al. Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* 2015;21:751–9.
- Siegel MB, He X, Hoadley KA, Hoyle A, Pearce JB, Garrett AL, et al. Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. *J Clin Invest* 2018;128:1371–83.
- Avigdor BE, Cimino-Mathews A, DeMarzo AM, Hicks JL, Shin J, Sukumar S, et al. Mutational profiles of breast cancer metastases from a rapid autopsy series reveal multiple evolutionary trajectories. *JCI Insight* 2017;2.
- Savas P, Teo ZL, Lefevre C, Flensburg C, Caramia F, Alsop K, et al. The subclonal architecture of metastatic breast cancer: results from a prospective community-based rapid autopsy program “CASCADE.” Ladanyi M, editor. *PLOS Med* 2016;13:e1002204.
- Markowitz SD. Cancer bypasses the lymph nodes. *Science* 2017;357:35–6.
- Wong SY, Hynes RO. Lymphatic or hematogenous dissemination: How does a metastatic tumor cell decide? *Cell Cycle* 2006;5:812–7.
- Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 2016;352:169–75.
- 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;534:47–54.
- Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* 2011;144:646–74.
- Sethi N, Kang Y. Unravelling the complexity of metastasis-molecular understanding and targeted therapies. *Nature Reviews Cancer* 2011;11:735–48.
- Vanharanta S, Massagué J. Origins of metastatic traits. *Cancer Cell* 2013;24:410–21.
- Singhi AD, Cimino-Mathews A, Jenkins RB, Lan F, Fink SR, Nassar H, et al. MYC gene amplification is often acquired in lethal distant breast cancer metastases of unamplified primary tumors. *Mod Pathol* 2012;25:378–87.
- Wolfer A, Ramaswamy S. MYC and metastasis. *Cancer Research* 2011;71:2034–7.
- Ivens D, Hoe AL, Podd TJ, Hamilton CR, Taylor I, Royle GT. Assessment of morbidity from complete axillary dissection. *Br J Cancer* 1992;66:136–8.
- Jatoi I, Benson JR, Toi M. De-escalation of axillary surgery in early breast cancer. *Lancet Oncol* 2016;17:e430–41.
- Chen AY, Halpern MT, Schrag NM, Stewart A, Leitch M, Ward E. Disparities and trends in sentinel lymph node biopsy among early-stage breast cancer patients (1998–2005). *J Natl Cancer Inst* 2008;100:462–74.