European Journal of Cancer 143 (2021) 147–157

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Understanding the organ tropism of metastatic breast cancer through the combination of liquid biopsy tools

Lorenzo Gerratana [a,b,c], Andrew A. Davis [a,d], Maurizio Polano [e], Qiang Zhang [a], Ami N. Shah [a], Chenyu Lin [a], Debora Basile [b,c], Giuseppe Toffoli [e], Firas Wehbe [f], Fabio Puglisi [b,c], Amir Behdad [g], Leonidas C. Platanias [a], William J. Gradishar [a], Massimo Cristofanilli [a,*]

[a] *Department of Medicine-Hematology and Oncology, Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*
[b] *Department of Medicine (DAME), University of Udine, Udine, UD, Italy*
[c] *Department of Medical Oncology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, PN, Italy*
[d] *Department of Medicine, Division of Oncology, Washington University School of Medicine in St. Louis, MO, USA*
[e] *Experimental and Clinical Pharmacology Unit, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, PN, Italy*
[f] *Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*
[g] *Department of Pathology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA*

Received 19 May 2020; received in revised form 1 November 2020; accepted 12 November 2020
Available online 8 December 2020

KEYWORDS
Precision medicine;
Organotropism;
Liquid biopsy;
Metastatic breast cancer;
Circulating tumour cell;
Circulating tumour DNA

Abstract ***Background:*** Liquid biopsy provides *real-time* data about prognosis and actionable mutations in metastatic breast cancer (MBC). The aim of this study was to explore the combination of circulating tumour DNA (ctDNA) analysis and circulating tumour cells (CTCs) enumeration in estimating target organs more susceptible to MBC involvement.
Methods: This retrospective study analysed 88 MBC patients characterised for both CTCs and ctDNA at baseline. CTCs were isolated through the CellSearch kit, while ctDNA was analysed using the Guardant360 NGS-based assay. Sites of disease were collected on the basis of imaging. Associations were explored both through uni- and multivariate logistic regression and Fisher's exact test and the random forest machine learning algorithm.
Results: After multivariate logistic regression, *ESR1* mutation was the only significant factor associated with liver metastases (OR 8.10), while *PIK3CA* was associated with lung localisations (OR 3.74). CTC enumeration was independently associated with bone metastases (OR 10.41) and *TP53* was associated with lymph node localisations (OR 2.98). The metastatic behaviour was further investigated through a random forest machine learning algorithm. Bone

* *Corresponding author*: Robert H. Lurie Comprehensive Cancer Center, Northwestern University, Feinberg School of Medicine, 710 North Fairbanks Court, Chicago, IL, 60611, USA.
E-mail address: massimo.cristofanilli@nm.org (M. Cristofanilli).

https://doi.org/10.1016/j.ejca.2020.11.005

involvement was described by CTC enumeration and alterations in *ESR1*, *GATA3*, *KIT*, *CDK4* and *ERBB2*, while subtype, CTC enumeration, inflammatory BC diagnosis, *ESR1* and *KIT* aberrations described liver metastases. *PIK3CA, MET, AR*, CTC enumeration and *TP53* were associated with lung organotropism. The model, moreover, showed that *AR, CCNE1, ESR1, MYC* and CTC enumeration were the main drivers in HR positive MBC metastatic pattern.

Conclusions: These results indicate that ctDNA and CTCs enumeration could provide useful insights regarding MBC organotropism, suggesting a possible role for future monitoring strategies that dynamically focus on high-risk organs defined by tumourbiology.

1. Background

Breast cancer (BC) is the most frequently diagnosed cancer and the leading cause of cancer death in women [1]. Although about 6–7% of newly diagnosed cases present as *de novo* metastatic disease, distant involvement usually occurs as a later event in approximately 30% of patients initially diagnosed with early-stage BC [2,3]. For this reason, knowing the potential organotropism of the disease could be critical for an effective personalised follow-up strategy [4].

The combined immunohistochemistry (IHC) and *in situ* hybridisation (ISH) diagnostic approach on the basis of hormone receptors (HRs) and human epidermal growth factor receptor 2 (HER2) expression and amplification are currently the most common clinical tools used for BC classification. It is recognised that each clinical subtype has different preferential sites of metastatic involvement. For example, hormone-receptor positive (HR positive) BC more commonly metastasises to the bone, a site that is less likely to be observed in triple-negative (TN) BC. Furthermore, HR-positive BC has a less pronounced organotropism for visceral organs as compared with HER2-positive BC and TNBC which often metastasises to the central nervous system (CNS), liver, and lungs [3,5]. There are currently no diagnostic tools with the ability to predict site of recurrences and facilitate monitoring and design of organ-specific interventions. Liquid biopsy provides *real-time* data in metastatic breast cancer (MBC) through the detection and isolation of circulating tumour DNA (ctDNA), circulating tumour cells (CTCs), exosomes, and other blood-based biomarkers for prognostic stratification and genomic characterisation.

Previous studies have been performed to utilise a liquid biopsy paradigm to better define risk stratification by site of disease using Circulating tumour Cells (CTCs) characterisation. On the basis of a DEPArray-based strategy, CTCs were classified as epithelial, epithelial to mesenchymal, mesenchymal and negative, highlighting a significant association between epithelial CTCs, bone and liver involvement and negative circulating cells and CNS [6]. These data suggest the potential to capture and describe the biological characteristics of the disease through liquid biopsy, with possible implications both in terms of the depth of characterisation achievable and in its scalability over time.

Machine learning (ML) is an application of artificial intelligence that provides systems that automatically learn on the basis of training datasets without being explicitly programmed. Through an ability to recognise and analyze patterns, ML algorithms are able to adaptively improve performance and define natural groups or latent parameters in data through generalising solutions. This potential can be applied to the clinical and biomedical field, where ML offers predictive models that are able to map highly heterogeneous data, even when relationships could not be determined due to complexity or lack of biological understanding [7,8]. Among the different ML algorithms, random forest is one of the most widespread and flexible. A random forest algorithm consists of a large number of individual decision trees that operate as an ensemble. Decision trees are a list of all possible alternative modalities that every feature in the model can have, each individual tree in the random forest model gives an independent classification prediction and the one with the highest votes is considered the most reliable [7,8]. The application of this approach to studying sites of metastatic recurrence in BC using clinical and genomic data is currently limited.

Based on these premises, the aim of this study was to explore the combination of clinical characteristics, ctDNA-detected aberrations, and CTC enumeration in estimating target organs more susceptible to MBC involvement. This study examined a novel ML modelling approach to describe the metastatic behaviour as an emerging propriety on the basis of clinical features and detectable gene alterations.

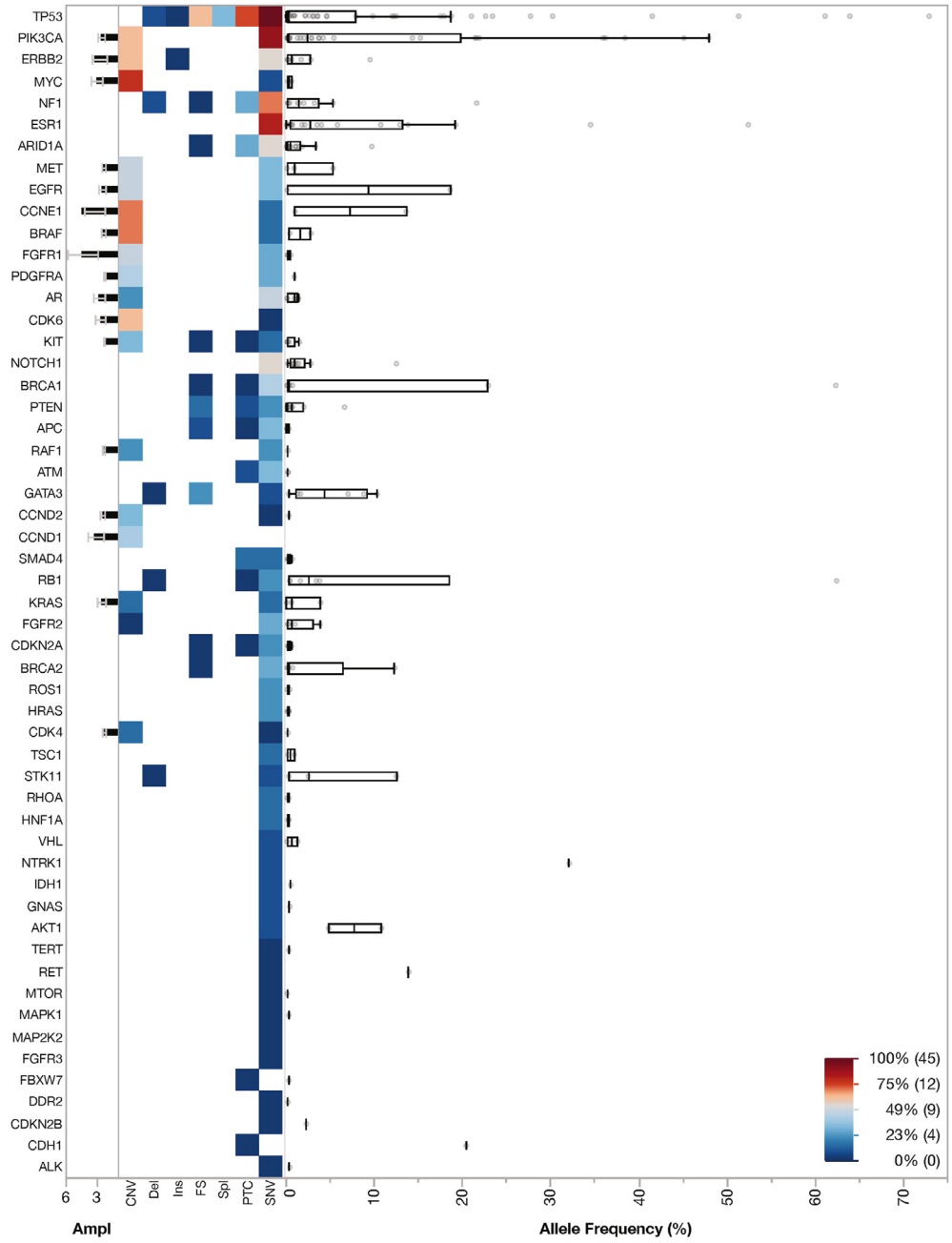

Fig. 1. **Landscape plot of all detected alterations in the analysed cohort**. Incidence of the single aberrations [copy number variations (CNV), deletion (Del), insertion (ins), splicing variants (Spl), premature termination codons (PTC) and single nucleotide variation (SNV)] are represented on the center, while the magnitude of the described CNVs is showed on the left and the mutant allele frequency (MAF) of each mutation on the right. TP53, PIK3CA, ERBB2, MYC, NF1, ESR1 ARID1A and MET were the most commonly altered genes (red = above the median), while PIK3CA, ERBB2, MYC and MET had also CNVs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2. Methods

2.1. Cohort design

The study analysed a cohort of 88 MBC patients treated and characterised for CTCs and ctDNA at the Robert H. Lurie Comprehensive Cancer Center at Northwestern University (Chicago, IL, USA) between 2016 and 2018. Patient enrolment was performed under the Investigator Initiated Trial (IIT) NU16B06. CTC and ctDNA collection were performed at baseline, prior to treatment start. No selection was made on the basis of treatment line.

The baseline staging was performed before treatment start concomitantly to liquid biopsy sampling. Imaging was chosen according to the investigators' common

practice [e.g. Computed Tomography (CT) Scan, Positron Emission Tomography (PET)]. Distant localisations were categorised based to the presence of specific organ involvement (e.g. liver involvement, yes vs. no) independently from other metastatic sites.

2.2. CTC detection and enumeration

CTC analysis was performed though the CellSearch™ immunomagnetic System (Menarini Silicon Biosystems, PA, USA). CellSave Stabilising Tubes were used for blood collection (10 ml of whole blood). Sample were processed via Celltracks Autoprep for EpCAM-based immunomagnetic sorting and subsequent characterisation for pan-cytokeratin (CK), fluorescent dye 4'-6-Diamidino-2-phenylindole (DAPI), and CD45 cells. The enriched and labelled samples were then reviewed via the Celltracks Analyzer II. CTCs were defined as CK/EpCAM/DAPI positive, CD45 negative. Patients with ≥5 CTC/7.5 ml of blood were defined as Stage $IV_{aggressive}$ as previously reported [9].

2.3. ctDNA sequencing

ctDNA samples were analysed using the Guardant360™ next-generation sequencing (NGS) commercial platform (Guardant Health, CA). Two 10-ml of whole blood were drawn for each patient using standard stabilising tubes (Streck, NE). ctDNA extraction (QIAGEN Inc., MD) was performed in a Clinical Laboratory Improvement Amendments (CLIA)-certified, College of American Pathologists (CAP)-accredited, clinical laboratory [10]. NGS was on the basis of a 74 gene panel using a single-molecule digital sequencing technology for somatic single nucleotide variants (SNVs), insertions/deletions (indels), gene fusions/rearrangements and copy number variations (CNVs) as previously described with reported analytical sensitivity and specificity >99% [10–12]. The range of genes examined in this particular cohort is represented in Fig. 1.

2.4. Statistical analysis

Clinical and pathological variables were reported using descriptive analyses. Categorical variables were reported as frequency distribution, whereas continuous variables were described according to the median and interquartile range (IQR). CTC enumeration was considered both as a continuous variable (nCTCs) and dichotomised according to the threshold of ≥5 CTC/7.5 ml (Stage IV aggressive versus Stage IV indolent) [9]. Differences in distribution of continuous variables such as nCTCs, number of detected alterations (NDA) and mutant allele frequency (MAF) of the highest clone across different metastatic sites were tested through the Mann–Whitney *U* test.

Table 1
Clinico-pathological characteristics of the analysed cohort.

	N	%
Age		
<50	35	39.77
≥50	53	60.23
IBC		
No	45	51.14
Yes	43	48.86
BC Subtype		
Luminal-like	35	43.21
HER2 positive	20	24.69
Triple Negative	26	32.10
Liver		
No	58	65.91
Yes	30	34.09
Lung		
No	55	62.50
Yes	33	37.50
Bone		
No	44	50.00
Yes	44	50.00
Lymph node		
No	46	52.27
Yes	42	47.73
CNS		
No	80	90.9
Yes	8	9.1
Serosa		
No	79	89.8
Yes	9	10.2
CTC enumeration		
Median	1	IQR 0–12.5
Stage $IV^{indolent}$		58 65.91
Stage $IV^{aggressive}$		30 34.09
MAF		
Median	2.6	IQR 0.4–15.2
NDA		
Median	4	IQR 2–7

The associations among CTCs, specific genomic alterations, and metastatic sites were explored through unconditional uni- and multivariate logistic regression inclusive of odds ratio (OR) and a 95% confidence interval (95% CI) computation. Two-sided Fisher's exact test was applied when statistically appropriate. To ensure model stability, gene variants were tested if adequately represented in the population (Supplementary Table 1).

Statistical analysis was conducted using the StataCorp 2016 Stata Statistical Software: Release 14.2 (College Station, Texas, USA).

2.5. Machine learning

Random Forrest was implemented on python using h2oai. The optimal hyperparameters were selected using grid search approach to increase the accuracy of specific model [13,14]. The data were split beforehand into 80% training and 20% test partitions. Model was developed with balance Class option activate for increase the

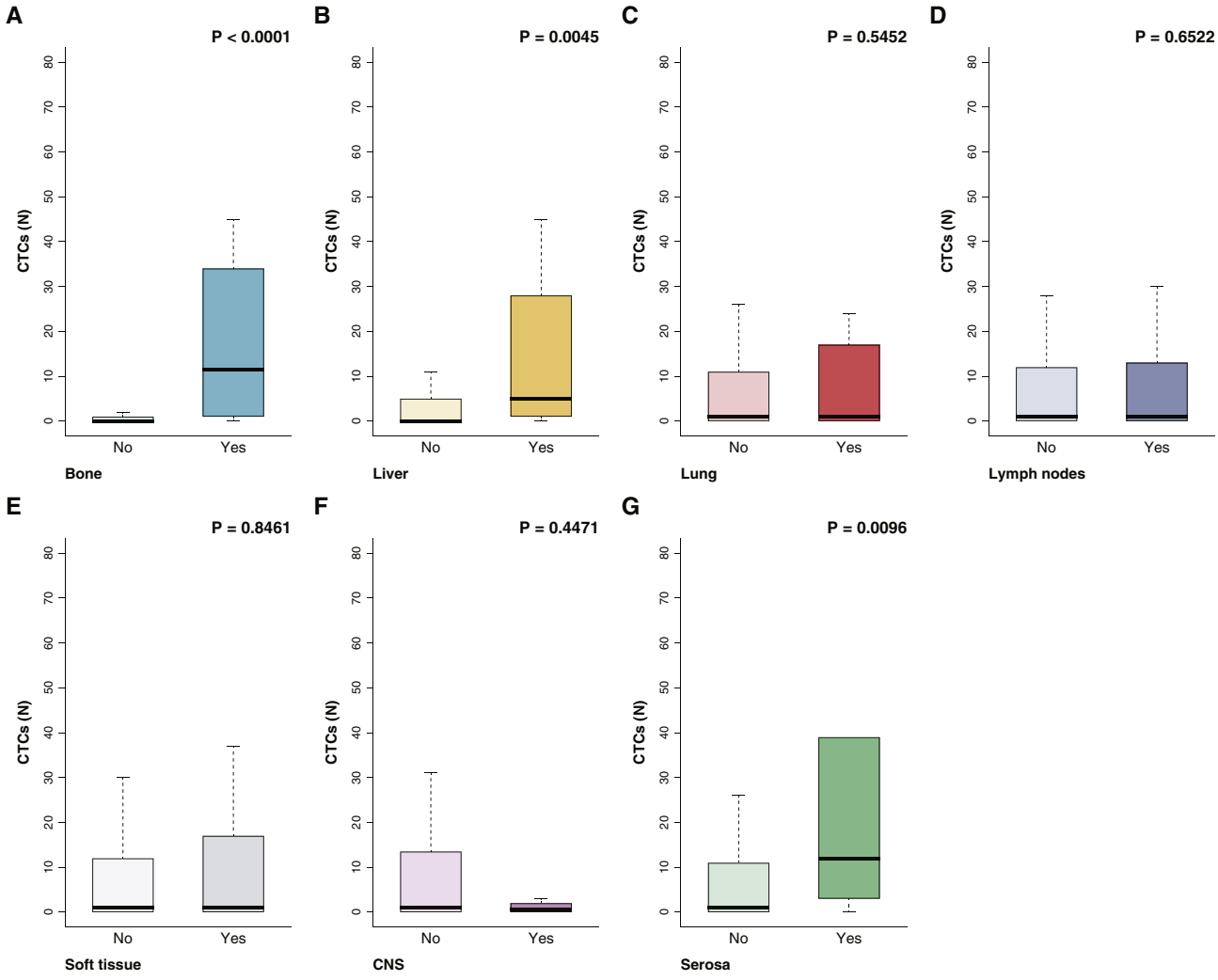

Fig. 2. **Box plot of the nCTCs distribution across different metastatic sites**. nCTCs was significantly higher in patients with bone metastasis (A), liver involvement (B) and serosal localisation (G). No significant associations were observed for the remaining metastatic sites. Significance was tested though the Mann–Whitney *U* test.

weight of minority class. All models were developed in a 10× 5-fold cross validation (CV) schema on the training partition. Performance was assessed in terms of accuracy (ACC). Feature importance reflects the role of each single feature within the whole model suggesting the putative relation from the data.

All the experiments were run on a 32-core Intel Core i7 workstation with 128GB of RAM running CentOS 7.5.

The models comprised all detected gene variants (Fig. 1), together with IBC diagnosis, CTC enumeration and BC subtype.

3. Results

3.1. Cohort characteristics and detected gene alterations

The cohort consisted of 88 metastatic breast cancer (MBC) patients with a median age at the first blood draw for ctDNA of 55years (range: 29–82years). Forty-three patients (49%) were diagnosed with inflammatory BC. Moreover, 43% had hormone receptor (HR) positive MBC, 32% had triple negative breast cancer (TNBC), and 25% had human epidermal growth factor receptor 2 (HER2)-positive MBC with 34% defined as stage IV $_{aggressive}$. The median number of prior treatment lines at baseline collection was 1 (IQR: 1–3), while the median number of metastatic sites was 3 (IQR: 1–3) with the most observed sites being lymph nodes (42%), lung (38%), bone (34%) and liver (34%) (Table 1).

Across the tested genes, *TP53, PIK3CA, ERBB2, MYC, NF1, ESR1 ARID1A* and *MET* were the most commonly observed ctDNA-detected aberrations (Fig. 1). Notably *PIK3CA, ERBB2, MYC* and *MET* were the most common copy number variations (CNVs) (Fig. 1).

The median number of detected alterations (NDA) per patient was 4 (IQR: 2–7), while the median mutant

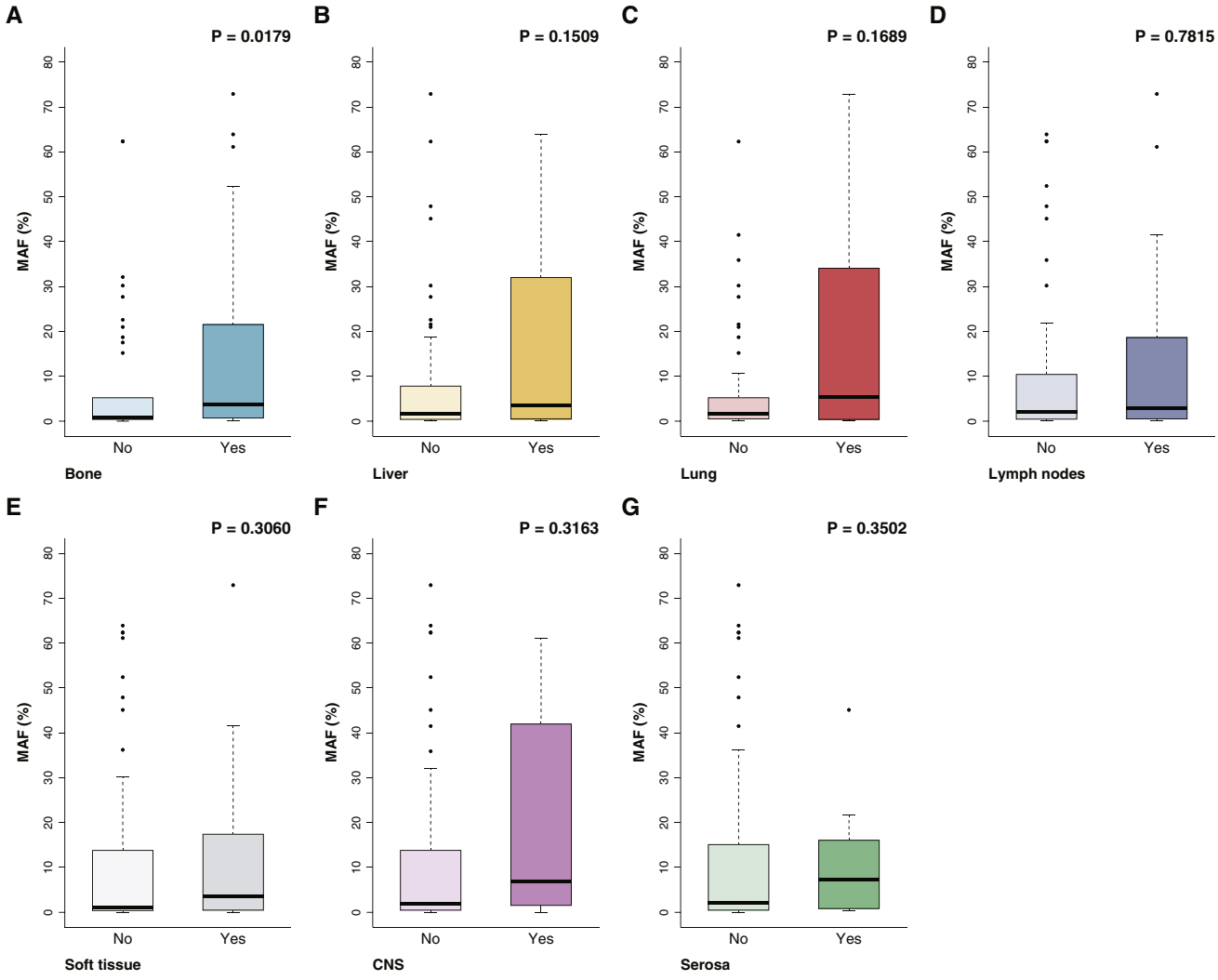

Fig. 3. **Box plot of the MAF distribution across different metastatic sites**. MAF was significantly higher in patients with bone metastasis (A). No significant associations were observed for the remaining metastatic sites. Significance was tested though the Mann–Whitney *U* test.

allele frequency (MAF) of the highest clone was 2.6% (IQR: 0.4–15.2%).

3.2. NDA, MAF and nCTCs across different metastatic sites

The association with different sites of distant involvement was tested for NDA, MAF, and nCTCs.

nCTCs was significantly higher in patients with bone (median nCTCs: 11.5, IQR: 1–34 versus median nCTCs: 0, IQR: 0–1 respectively in bone yes versus no) ($P < 0.0001$) (Fig. 2A), liver (median nCTCs: 5, IQR 1–28 and median CTCs: 0, IQR 0–5 respectively in liver yes versus no) ($P = 0.0045$) (Fig. 2B) and serosal involvement (median nCTCs: 12, IQR: 3–39 and median CTCs: 1, IQR: 0–11 respectively in serosa yes versus no $P = 0.0096$) (Fig. 2G).

Higher MAF was observed in patients with bone involvement (median MAF: 3.8, IQR 0.6–21.6 and median MAF: 0.8, IQR 0.3–5.3 respectively in bone yes versus no $P = 0.0179$) (Fig. 3A).

Significantly higher NDA was observed in patients with liver metastases (median NDA: 6.5, IQR 2–10 and median NDA: 4, IQR 2–6 respectively in liver yes versus no $P = 0.0424$) (Fig. 4B).

No significant associations were observed for the remaining metastatic sites (Figs. 2–4).

3.3. Associations between ctDNA-detected gene alterations and metastatic sites

The association between the detected gene alterations (Fig. 1), CTC enumeration, clinical characteristics and metastatic sites was tested through univariate analysis (Supplemental Table 1) and confirmed in the resulting multivariate models. *ESR1* mutations was the only significant factor associated with liver metastases in multivariate analysis (OR 8.10, $P = 0.025$), while

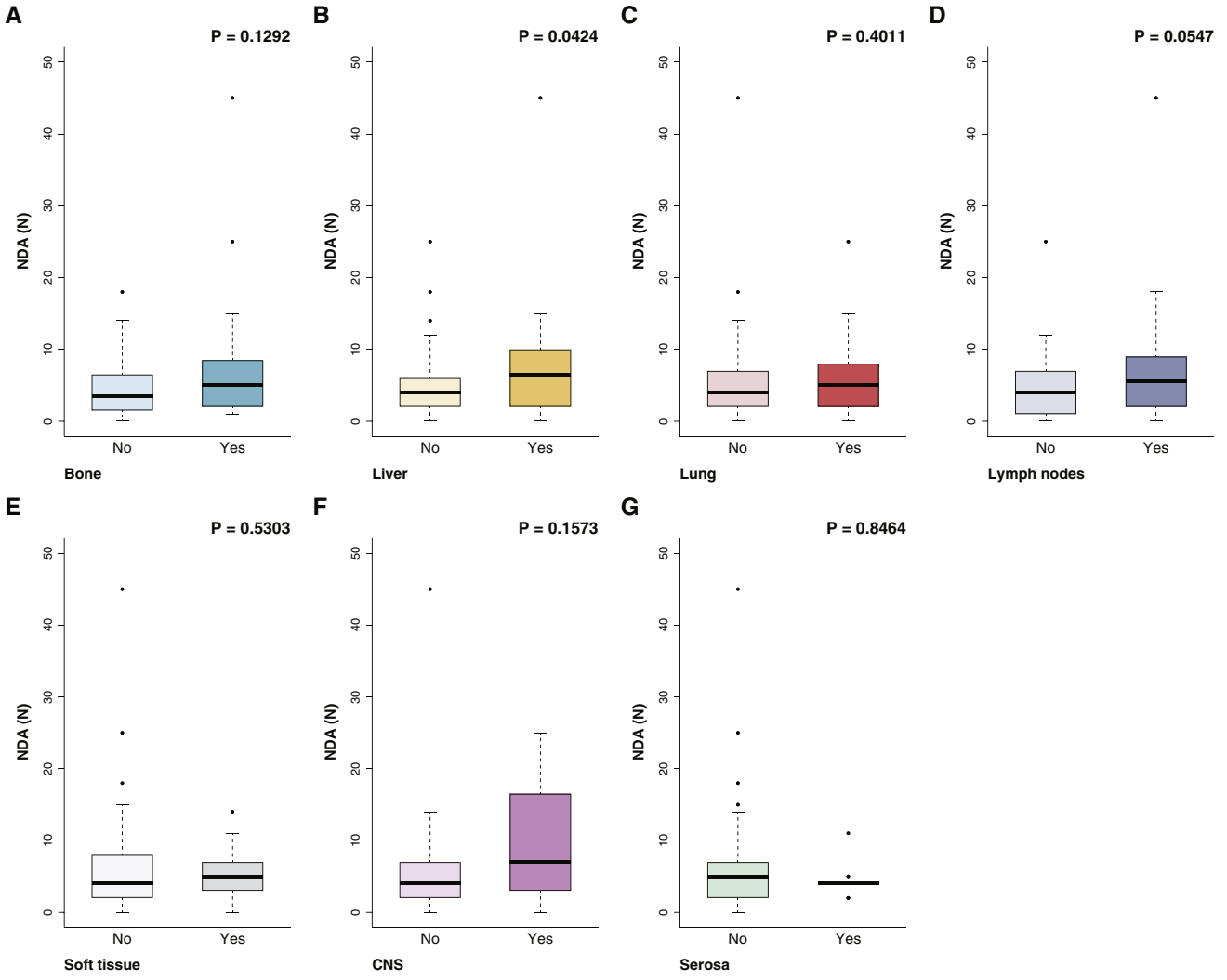

Fig. 4. **Box plot of the NDA distribution across different metastatic sites**. NDA was significantly higher in patients with liver involvement (B). A numerical difference was observed for lymph node localisations (D). No significant associations were observed for the remaining metastatic sites. Significance was tested though the Mann–Whitney *U* test.

PIK3CA was associated with lung localisations (OR 3.74, P = 0.010) (Table 2). All HR positive MBC patients with *ESR1* mutations had bone metastases (P = 0.022), while Stage $IV_{aggressive}$ was independently associated with bone metastases (OR 10.41, P = <0.001) (Table 2) [9]. *TP53 was* associated with lymph node localisations (OR 2.98, P = 0.032), while *CCND1* alterations were associated with serosal involvement (OR 24.58, P = 0.003) (Table 2). Notably, TNBC was associated with soft tissue spreading (OR 3.42, P = 0.040) (Table 2).

Among patients with liver metastases, eight individuals had a single *ESR1* variant (D538G, E380Q, L536P, L536R, Y537N or Y537S), while three showed concomitant variants (D538G/Y537N/Y537, E380Q/F461I/P535R/G442R, or Y537S/Y537N/D538G respectively). Amplification was the most observed *PIK3CA* alteration observed in patients with lung metastases, both as a single alteration and concomitantly with other *PIK3CA* variants (Amplification/E81A, E542K/E726K/Amplification or H1047R/Amplification), while three patients showed concomitant single nucleotide variations (SNVs) (E542K/E726K, E545K/D1017H and H1047R/N426S), the remaining detected *PIK3CA* variants were exclusive (E542K, E545K, H1047R, N345K or Q546K). A high number of concomitant *TP53* variants was observed in patients with lymph nodes involvement (A159V/R248W, Exon 5 Deletion/H193L/R280G, G245S/D281Y/K101*/R181P, K132R/R248W, P151A/I50fs, Q167*/H179D, R196*/M237I, R248Q/Y163N/P177L, R273H/M246V or R273H/R209fs). *CCND1* amplification was the only variant observed in patients with serosal localisations.

3.4. Contribution of clinical and liquid biopsy-derived features across metastatic sites through an exploratory machine learning approach

The potential of combining clinical and liquid biopsy-derived features to describe the metastatic behaviour of patients with MBC was investigated. Bone involvement was mainly described by nCTC, and genomic abnormalities in *ESR1, GATA3, KIT, CDK4* and *ERBB2* (ACC: 0.743 ± 0.240) (Fig. 5A). Disease subtype (i.e. HR positive, HER2 positive and TNBC), nCTC, inflammatory BC diagnosis, and aberrations in *ESR1* and *KIT* highly contributed in describing liver metastases (ACC: 0.852 ± 0.184) (Fig. 5B), while *PIK3CA, MET, AR*, nCTC and *TP53* were associated with lung organotropism (ACC: 0.502 ± 0.268) (Fig. 5C).

Gene alterations in *NF1 CDK6, MYC, MET* and *CCNE1* and nCTCs were linked to nodal involvement (ACC: 0.558 ± 0.39) (Fig. 5D), while *ERBB2*, *FGFR1* and nCTC were linked to soft tissue involvement (ACC: 0.561 ± 0.32) (Fig. 5E). *CCNE1, FGFR1, TP53, MYC* and *ERBB2* were features linked to CNS localisations (ACC: 0.450 ± 0.04) (Fig. 5F).

Accuracy coefficients relative to all models across metastatic sites are reported in Table 3.

The model was then implemented to investigate the overall metastatic pattern across MBC subtypes by considering not only single metastatic sites but also their combination. *AR, CCNE1, ESR1, MYC* and nCTC were the main drivers in HR positive MBC (mean per class error 0.467, cross validation accuracy 0.4772−0.0321) (Fig. 6). Modeling was not reliable in the HER2 positive and TNBC subgroups due to insufficient sample size (data not shown).

Table 2
Multivariate logistic regression models across different metastatic sites. Liver metastases were associated with ESR1 mutations, while PIK3CA was associated with lung localisations. Bone metastases were associated with Stage IV aggressive, ESR1 mutations were not inserted in the model due to the high concordance of patients with an ESR1 alteration and bone metastases (Fisher's exact test P = 0.022). Lymph node localisations were associated with *TP53* alterations. *CCND1* alterations were associated with serosal involvement.

		Odds Ratio	95% C.I.		P Value
Liver					
Subtype	HR positive	1			
	HER2 positive	0.14	0.01	1.27	0.080
	TNBC	0.84	0.24	2.94	0.788
CTC Enumeration	Stage $IV_{Indolent}$	1			
	Stage $IV_{Aggressive}$	1.46	0.43	4.92	0.539
ARID1A	Wild type	1			
	Altered	1.80	0.21	15.69	0.596
AR	Wild type	1			
	Altered	1.74	0.32	9.38	0.518
ESR1	Wild type	1			
	Altered	8.10	1.30	50.31	0.025
Bone*					
Subtype	HR positive	1			
	HER2 positive	0.26	0.07	1.07	0.063
	TNBC	0.31	0.08	1.26	0.103
CTC Enumeration	Stage $IV_{Indolent}$	1			
	Stage $IV_{Aggressive}$	10.41	2.80	38.67	<0.001
MYC	Wild type	1			
	Altered	3.35	0.78	14.45	0.104
TP53	Wild type	1			
	Altered	0.49	0.14	1.76	0.277
Lung					
Subtype	HR positive				
	HER2 positive	0.80	0.22	2.95	0.741
	TNBC	2.53	0.83	7.72	0.104
PIK3CA	Wild type	1			
	Altered	3.74	1.37	10.20	0.010
Lymph nodes					
TP53	Wild type	1			
	Altered	2.98	1.10	8.12	0.032
CCNE1	Wild type	1			
	Altered	1.65	0.40	6.80	0.488
FGFR1	Wild type	1			
	Altered	3.59	0.79	16.17	0.097
RAF1	Wild type	1			
	Altered	8.82	0.88	88.68	0.064
Soft tissue					
Subtype	HR positive	1			
	HER2 positive	2.55	0.76	8.49	0.128
	TNBC	3.42	1.06	11.05	0.040
TP53	Wild type	1			
	Altered	2.23	0.75	6.63	0.150
Serosa					
CTC Enumeration	Stage $IV_{Indolent}$	1			
	Stage $IV_{Aggressive}$	3.24	0.45	23.21	0.241
CCND1	Wild type	1			
	Altered	24.58	2.89	209.04	0.003
ESR1	Wild type	1			
	Altered	2.73	0.39	19.14	0.313
TP53	Wild type	1			
	Altered	0.20	0.03	1.39	0.105

4. Discussion

Although metastasis has been well characterised to drive clinical prognosis, the underlying biology needs to be further elucidated. The present study analysed a cohort of MBC patients characterised for both CTCs and ctDNA with the aim of capturing the *real-time* biological features of the disease according to different sites of metastasis.

First, the study showed that specific metastatic sites were associated with higher nCTCs NDA, and MAF, suggesting that potential differences in biology linked to specific metastatic sites resulted in a detectable read-out though liquid biopsy.

Second, the study examined the contribution of nCTCs and genomic alterations though uni- and multivariate logistic regression analysis, highlighting the preeminent role of *ESR1* and CTCs enumeration with respect to liver and bone metastases. These findings were in line with the observed association of bone metastases with higher CTCs or Stage $IV_{aggressive}$ and the strong association with the absolute and relative levels of CTCs expressing epithelial markers after EpCam independent enrichment [6,9,15].

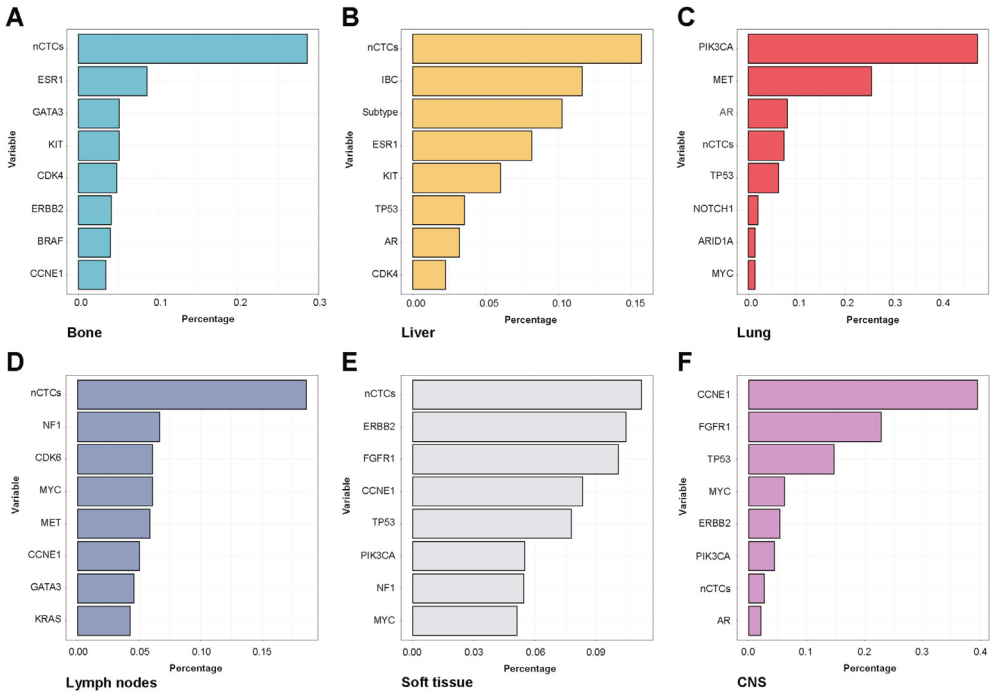

Fig. 5. **Relative importance of biomarker across different metastatic sites**. Histograms show VIMP (Variable importance. High importance values indicate variables with predictive ability, whereas zero or negative values identify non-predictive variables to be filtered.

Table 3
Random forest models performances. Accuracy coefficient relative to all models for each metastatic site.

Response	ACC (mean ± sd)	ACC test
Liver	0.853 ± 0.183	0.7714
Bone	0.743 ± 0.240	0.657
Lung	0.502 ± 0.268	0.799
CNS	0.450 ± 0.04;	0.75
Soft tissue	0.561 ± 0.32	0.44
Lymph Nodes	0.558 ± 0.39	0.736

The constant exposure to endocrine therapy (ET) agents such as aromatase inhibitors (AIs) and selective estrogen modulators and degraders such as Tamoxifen and Fulvestrant, can induce the onset or the selection of genetic and epigenetic alterations of the *ESR1* gene, such as activating mutations in the ligand-binding domain or increased methylation of the promoter [16–19]. These alterations confer not only treatment resistance, but can potentially change cell phenotype through changes in the chromatin recruitment transcriptional network, which results in neomorphic properties (i.e. a novel gene function or pattern of gene expression) [20–22]. This is a crucial aspect, since it suggests how treatment resistance is linked not only to target-dependent alterations, but how resistance also affects the overall phenotype and consequently the disease's clinical behaviour.

Third, the interplay among ctDNA, CTCs and metastatic sites, was also explored though ML. Intriguingly, *ESR1* was a recurrent factor across the different random forest models, together with other potential factors, such as aberrations in *PIK3CA*, *GATA3*, *ERBB2*, *KIT*, *FGFR1*, *MYC*, *MET* and *AR*. Moreover, CTCs were also highlighted as an important factor in all the

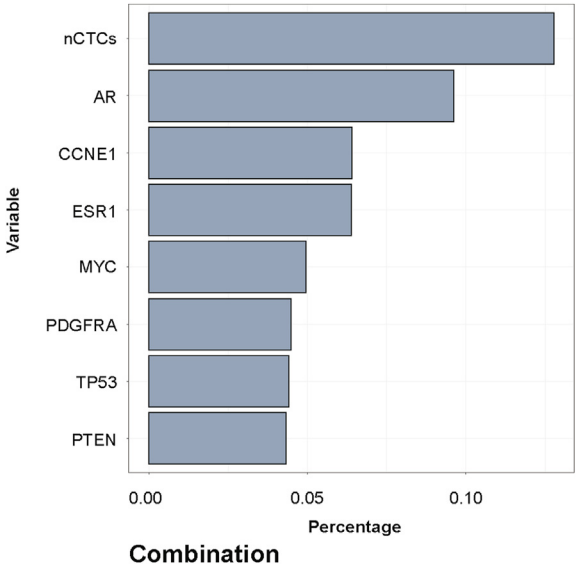

Fig. 6. **Factors linked to the overall metastatic pattern in HR positive MBC**. The model considered not only single metastatic sites but also their combination. Large importance values indicate variables with predictive ability, whereas zero or negative values identify non-predictive variables to be filtered.

investigated sites, with the exception of the CNS in which CTCs appeared to contribute less in our modeling. We previously reported the interplay between a higher nCTCs and genomic alterations in *ESR1, GATA3, CDH1*, and *CCND1*. The independent validation of these results with respect to *ESR1* and *GATA3* in the present study, further supports the relation between CTCs and distinctive biological features linked to *ESR1* [23].

In addition to describing the biology of MBC, the present study could provide initial clinical insights on MBC management. If validated, these results may define a potential role of a careful monitoring strategy with dynamic assessments using liquid biopsy. While baseline characteristics could detect high-risk patients, who may require a more intensive approach, the emergence of new gene alterations could identify patients more likely to develop certain disease localisations, suggesting a potential restaging strategy with the highest pre-test probability.

The application of ML to cancer genomics is a novel, yet promising, approach that is starting to show interesting implications in oncology. An algorithm trained on 7791 tumours prospectively sequenced in a cohort of patients with advanced cancer was capable to predict the correct tumour type in 5748 of the 7791 patients (73.8%) in the training set as well as 8623 of 11 644 patients (74.1%) in an independent cohort [24].

This approach yield to accurate predictions in 45 of 60 cases (75.0%) when applied to ctDNA samples [24]. Applying this method prospectively to patients under active care enabled genome-directed reassessment of diagnosis in 2 patients initially presumed to have metastatic breast cancer, leading to the selection of more appropriate treatments, which elicited clinical responses [24]. The present study, while still exploratory, demonstrates a novel liquid-biopsy based approach for integrating clinical and blood-based tumour data focused not only on relapse risk stratification, but also to define distant sites of disease more prone to metastatic involvement. This new concept could pave the way to new studies focused on liquid biopsy-assisted follow-up strategies in Early Breast Cancer.

The present study has several limitations. First, the relatively small sample size may have limited the ML and logistic regression performance, especially in underrepresented subgroups (e.g. CNS involvement). Second, there was a particularly high incidence of IBC patients in the cohort, which reflected the academic referral center bias where the data were collected. Therefore, there may be some limits to the generalisability of the findings. Third, the single time point of analysis warrants further studies focussed both on validation and on the implementation of a longitudinal model capable of capturing clonal evolution and the resulting changes in organotropism.

5. Conclusions

This novel approach suggests that integrating both CTC enumeration and genomic characterisation by ctDNA could help define sites of metastasis, paving the way for a more tailored monitoring and therapeutic approach.

Author contributions

Study concept and design: LG, AAD, MP, QZ, FP, MC; Data acquisition: LG, AAD, ANS, CL, FW, AB, LCP, WJG, MC; Quality control of data and algorithms: LG, AAD, MP, QZ, FP, MC; Statistical analysis: LG,MP; Interpretation of data, approval and editing of the manuscript: all authors.

Funding

Lynn Sage Cancer Research Foundation, OncoSET Precision Medicine Program, and REDCap support was funded in part by a Clinical and Translational Science Award (CTSA) grant from the National Institutes of Health UL1TR001422.

Conflict of interest statement

AAD reports non-financial support from Menarini Silicon Biosystems, outside the submitted work; LG reports non-financial support from Menarini Silicon Biosystems, personal fees from Lilly, outside the submitted work; MC reports grants from National Institutes of Health, during the conduct of the study; personal fees from Pfizer, personal fees from Merus, personal fees from Novartis, personal fees from CytoDyn, outside the submitted work; FP reports grants from AstraZeneca, grants from Eisai, outside the submitted work; all other authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2020.11.005.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics. CA Cancer J Clin 2019;69:7–34. https://doi.org/10.3322/caac.21551.

[2] Bonotto M, Gerratana L, Poletto E, Driol P, Giangreco M, Russo S, et al. Measures of outcome in metastatic breast cancer: insights from a real-world scenario. Oncologist 2014;19:608–15. https://doi.org/10.1634/theoncologist.2014-0002.

[3] Kennecke H, Yerushalmi R, Woods R, Cheang MCU, Voduc D, Speers CH, et al. Metastatic behavior of breast cancer subtypes. J Clin Oncol 2010;28:3271–7. https://doi.org/10.1200/JCO.2009.25.9820.

[4] Puglisi F, Fontanella C, Numico G, Sini V, Evangelista L, Monetti F, et al. Follow-up of patients with early breast cancer: is

it time to rewrite the story? Crit Rev Oncol Hematol 2014;91:130–41. https://doi.org/10.1016/j.critrevonc.2014.03.001.

[5] Gerratana L, Fanotto V, Bonotto M, Bolzonello S, Minisini AM, Fasola G, et al. Pattern of metastasis and outcome in patients with breast cancer. Clin Exp Metastasis 2015;32:125–33. https://doi.org/10.1007/s10585-015-9697-2.

[6] Bulfoni M, Gerratana L, Del Ben F, Marzinotto S, Sorrentino M, Turetta M, et al. In patients with metastatic breast cancer the identification of circulating tumor cells in epithelial-to-mesenchymal transition is associated with a poor prognosis. Breast Cancer Res 2016;18:30. https://doi.org/10.1186/s13058-016-0687-3.

[7] Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. J Med Syst 2002;26:445–63. https://doi.org/10.1023/A:1016409317640.

[8] Forman G, Cohen I. In: Boulicaut J-F, Esposito F, Giannotti F, Pedreschi D, editors. Learning from little: comparison of classifiers given little training BT - knowledge discovery in databases: PKDD 2004. Berlin, Heidelberg: Springer Berlin Heidelberg; 2004. p. 161–72.

[9] Cristofanilli M, Pierga J-Y, Reuben J, Rademaker A, Davis AA, Peeters DJ, et al. The clinical use of circulating tumor cells (CTCs) enumeration for staging of metastatic breast cancer (MBC): international expert consensus paper. Crit Rev Oncol Hematol 2019;134:39–45. https://doi.org/10.1016/j.critrevonc.2018.12.004.

[10] Lanman RB, Mortimer SA, Zill OA, Sebisanovic D, Lopez R, Blau S, et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. PloS One 2015;10:e0140712. https://doi.org/10.1371/journal.pone.0140712.

[11] Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res 2017;45:D777–83. https://doi.org/10.1093/nar/gkw1121.

[12] Zill OA, Banks KC, Fairclough SR, Mortimer SA, Vowles JV, Mokhtari R, et al. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. Clin Cancer Res 2018;24:3528–38. https://doi.org/10.1158/1078-0432.CCR-17-3837.

[13] Huang BFF, Boutros PC. The parameter sensitivity of random forests. BMC Bioinf 2016;17:331. https://doi.org/10.1186/s12859-016-1228-x.

[14] Ko J, Baldassano SN, Loh P-L, Kording K, Litt B, Issadore D. Machine learning to detect signatures of disease in liquid biopsies – a user's guide. Lab Chip 2018;18:395–405. https://doi.org/10.1039/C7LC00955K.

[15] Cristofanilli M, Budd GT, Ellis MJ, Stopeck A, Matera J, Miller MC, et al. Circulating tumor cells, disease progression, and survival in metastatic breast cancer. N Engl J Med 2004;351:781–91. https://doi.org/10.1056/NEJMoa040766.

[16] Fribbens C, O'Leary B, Kilburn L, Hrebien S, Garcia-Murillas I, Beaney M, et al. Plasma ESR1 Mutations and the treatment of estrogen receptor-Positive advanced breast cancer. J Clin Oncol 2016;34:2961–8. https://doi.org/10.1200/JCO.2016.67.3061.

[17] O'Leary B, Hrebien S, Morden JP, Beaney M, Fribbens C, Huang X, et al. Early circulating tumor DNA dynamics and clonal selection with palbociclib and fulvestrant for breast cancer. Nat Commun 2018;9:896. https://doi.org/10.1038/s41467-018-03215-x.

[18] O'Leary B, Cutts RJ, Liu Y, Hrebien S, Huang X, Fenwick K, et al. The genetic landscape and clonal evolution of breast cancer resistance to palbociclib plus fulvestrant in the PALOMA-3 trial. Cancer Discov 2018;8:1390–403. https://doi.org/10.1158/2159-8290.CD-18-0264.

[19] Gerratana L, Basile D, Franzoni A, Allegri L, Viotto D, Corvaja C, et al. Plasma-based longitudinal evaluation of ESR1 epigenetic status in hormone receptor-positive HER2-negative metastatic breast cancer. Front Oncol 2020;10. https://doi.org/10.3389/fonc.2020.550185.

[20] O'Leary B, Lira M, Huang S, Deng S, Xie T, Kinong J, et al. Longitudinal ctDNA sequencing using an expanded genomic panel in the PALOMA3 trial of palbociclib plus fulvestrant. SABCS; 2018. Abstract No. PD2-02.

[21] Stone A, Zotenko E, Locke WJ, Korbie D, Millar EKA, Pidsley R, et al. DNA methylation of oestrogen-regulated enhancers defines endocrine sensitivity in breast cancer. Nat Commun 2015;6:7758. https://doi.org/10.1038/ncomms8758.

[22] McDonnell DP, Norris JD, yi Chang C. Neomorphic ERα mutations drive progression in breast cancer and present a challenge for new drug discovery. Canc Cell 2018;33:153–5. https://doi.org/10.1016/j.ccell.2018.01.014.

[23] Davis AA, Zhang Q, Gerratana L, Shah AN, Zhan Y, Qiang W, et al. Association of a novel circulating tumor DNA next-generating sequencing platform with circulating tumor cells (CTCs) and CTC clusters in metastatic breast cancer. Breast Cancer Res 2019;21:137. https://doi.org/10.1186/s13058-019-1229-6.

[24] Penson A, Camacho N, Zheng Y, Varghese AM, Al-Ahmadie H, Razavi P, et al. Development of genome-derived tumor type prediction to inform clinical cancer care. JAMA Oncol 2019. https://doi.org/10.1001/jamaoncol.2019.3985.