



## ARTICLE

## Molecular Diagnostics

## Multi-protein spatial signatures in ductal carcinoma in situ (DCIS) of breast

Sunil S. Badve<sup>1</sup>, Sanghee Cho<sup>2</sup>, Yesim Gökmen-Polar<sup>1</sup>, Yunxia Sui<sup>2</sup>, Chrystal Chadwick<sup>2</sup>, Elizabeth McDonough<sup>2</sup>, Anup Sood<sup>2</sup>, Marian Taylor<sup>3</sup>, Maria Zavodszky<sup>2</sup>, Puay Hoon Tan<sup>4</sup>, Michael Gerdes<sup>2</sup>, Adrian L. Harris<sup>3</sup> and Fiona Ginty<sup>2</sup>

**BACKGROUND:** There is limited knowledge about DCIS cellular composition and relationship with breast cancer events (BCE).

**METHODS:** Immunofluorescence multiplexing (MxIF) was used to image and quantify 32 cellular biomarkers in FFPE DCIS tissue microarrays. Over 75,000 DCIS cells from 51 patients (median 9 years follow-up for non-BCE cases) were analysed for profiles predictive of BCE. K-means clustering was used to evaluate cellular co-expression of epithelial markers with ER and HER2.

**RESULTS:** Only ER, PR and HER2 significantly correlated with BCE. Cluster analysis identified 6 distinct cell groups with different levels of ER, Her2, cMET and SLC7A5. Clusters 1 and 3 were not significant. Clusters 2 and 4 (high ER/low HER2 and SLC7A5/mixed cMET) significantly correlated with low BCE risk ( $P = 0.001$  and  $P = 0.034$ ), while cluster 6 (high HER2/low ER, cMET and SLC7A5) correlated with increased risk ( $P = 0.018$ ). Cluster 5 (similar to cluster 6, except high SLC7A5) trended towards significance ( $P = 0.072$ ). A continuous expression score (Escore) based on these 4 clusters predicted likelihood of BCE (AUC = 0.79, log-rank test  $P = 5E-05$ ; LOOCV AUC = 0.74, log-rank test  $P = 0.006$ ).

**CONCLUSION:** Multiplexed spatial analysis of limited tissue is a novel method for biomarker analysis and predicting BCEs. Further validation of Escore is needed in a larger cohort.

*British Journal of Cancer* (2021) 124:1150–1159; <https://doi.org/10.1038/s41416-020-01216-6>

## BACKGROUND

The successful implementation of the breast screening program in developed countries has resulted in the identification of a large number of putative precursor lesions of invasive carcinoma. Ductal carcinoma in situ (DCIS) is a non-obligate precursor lesion that is managed aggressively. Most patients get treated with surgery followed by post-operative radiation therapy. These therapies have been documented to decrease the incidence of recurrence and development of invasive cancer. In addition, the UK/ANZ DCIS trial and the NSABP B-24 clinical trials further demonstrated a significant reduction in frequency of DCIS recurrence by the addition of endocrine therapy with resultant recurrence rates below 10%.<sup>1–3</sup>

DCIS, if left untreated, will progress to invasive carcinoma in only around 20–50% of patients.<sup>4–6</sup> This has led to significant concerns regarding overtreatment of patients. Currently, there are many trials that are enrolling patients with DCIS for non-surgical management based on histological features of DCIS. Low risk DCIS cases are being enrolled in the LORIS trial in the United Kingdom, LORD trial in Europe and LARRIKIN trial in Australia for non-surgical management by active surveillance.<sup>7–9</sup> This is similar to the COMET (comparing operative to monitoring and endocrine therapy for low risk DCIS) trial in the USA.<sup>10</sup> Inclusion criteria of the COMET Trial include women 40 years or older with screen-detected calcification associated with histologically confirmed low-grade (LG) or intermediate-grade (IG) DCIS. The LORIS Trial

and LORD Trial also study screen-detected calcification associated with LG or IG DCIS. However, there are some differences in age of eligibility, whether patients with bilateral or multicentric disease, prior breast disease or mantle radiation can be enrolled. The presence of comedo-necrosis is an important exclusion criterion in some of these trials. These histological features are subjective and there is poor inter-observer agreement due to intratumoural heterogeneity.<sup>11–13</sup> A recent survey of more than 30 international recognised breast pathologists documented marked variability in definition of comedo-necrosis.<sup>14</sup> This subjectivity will significantly impact patient enrolment and final study results. There is a clear need for better understanding the biology of DCIS and the pathways leading to (or associated with) progression.

A number of tools have been used for the prognostication of DCIS. These include histological features, single marker, as well as multi-marker panels for immunohistochemistry assays, in addition to multiplex RT-qPCR for mRNAs. Analysis of the 12-year follow-up data of the Eastern Cooperative Oncology Group (ECOG) E 5194 trial has confirmed the role of histological features (high grade) in predicting likelihood of recurrence.<sup>15</sup> This was also confirmed in analysis of 57,222 DCIS cases from the SEER database by Sagara et al.<sup>16</sup> The expression of oestrogen and progesterone receptors (ER and PR) is also associated with decreased risk of recurrence. In contrast, proliferation markers such as Ki67 are associated with a higher risk (reviewed in Groen et al.<sup>9</sup>). The Tlsty group has analysed the expression of p16, COX2 and Ki67 to identify an IHC

<sup>1</sup>Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA; <sup>2</sup>GE Research, Niskayuna, NY 12309, USA; <sup>3</sup>Department of Oncology, Cancer and Haematology Centre, Oxford University, Oxford OX37LJ, UK and <sup>4</sup>Department of Pathology, Singapore General Hospital, Singapore, Singapore  
Correspondence: Sunil S. Badve (sbadve@iupui.edu)

Received: 21 January 2020 Revised: 10 September 2020 Accepted: 25 November 2020

Published online: 7 January 2021

based predictor for the likelihood of recurrence.<sup>17</sup> In a multi-variable model, DCIS lesions that were p16<sup>+</sup>/COX2<sup>+</sup>/Ki67<sup>+</sup> or those detected by palpation were statistically significantly associated with subsequent invasive cancer.<sup>17</sup> Based on these initial analyses, they have identified a panel of IHC biomarkers (PR, HER2, Ki67, COX2, p16/INK4A, FOXA1 and SIAH2), which is commercially available through Prelude's CLIA-approved lab as DCIS on RT™ (Decision score (DS)). In multivariable analysis, DS, but not nuclear grade, correlated with the benefit of radiotherapy in the SweDCIS cohort.<sup>18</sup> In collaboration with Genomic Health Inc, we have described a 12-gene signature for DCIS.<sup>19</sup> This signature is based on mRNA levels of progesterone (*PgR*), Glutathione S-Transferase Mu 1 (*GSTM1*) and 5 proliferation genes (*MKI67*, *STK15*, *BIRC5*, *CCNB1*, and *MYBL2*). This has been further validated in independent studies, including the Toronto cohort.<sup>20</sup> We have also recently used multiplexed immunofluorescence imaging (MxIF) to study the degree of protein heterogeneity in DCIS.<sup>13</sup> This study documented marked heterogeneity at a single cell level in DCIS. One of the limitations of the study was the lack of clinical follow-up data, which limited understanding the impact of the heterogeneity in DCIS.

The current exploratory study was designed to further understand the impact of a large number of biomarkers previously reported to be associated with second breast cancer event (BCE) of DCIS using a cohort of well annotated DCIS cases with follow-up data. These included ER, PR, HER2, Ki67, p53, COX2, and CD10 which have been well described in DCIS. We further investigated the HER pathway by analysing EGFR, HER4 and PTEN. We also investigated the role of cancer stem cells (ALDH1 and CD44v6),<sup>21</sup> and proteins implicated in progression (p21, VEGFR2, cMET, CDCP1, HTF9C/TRMT2A, and CEACAM5)<sup>22–26</sup> and resistance to therapy (ABCB1, ABCG2, MRP4, MRP5, SLC7A5).<sup>27–30</sup> DAPI was used to identify the nuclei and together with NaKATPase, pan-cytokeratin and S6 used for epithelial cell segmentation. Multiplexed immunofluorescence and single cell analysis<sup>13,31</sup> was used to analyse the (co-) expression of these markers in a single FFPE TMA sections.

## METHODS

### Clinical Cohort

IRB permissions were obtained from Oxford University (for the entire study) and waiver of IRB from Indiana University. Identified DCIS cases were selected from the archives at Oxford University/Radcliffe General Hospital. The selection criteria were as follows: (1) they were excision specimens; (2) patients did not have invasive (or microinvasive) cancer; (3) patients did not have any prior therapy; (4) patients did not have prior history of breast cancer. The patients were diagnosed between 1986 and 2004 and selected on the basis of a long follow-up period, with a median of 8 years (range 1–17 years). This was to allow evaluation of the long natural history of DCIS biology. An initial histopathological review was performed to confirm the diagnosis of DCIS, which was confirmed independently by two pathologists.

Of the 135 patients, 62 were excluded due to insufficient clinical information or poor data quality, leaving 73 patients with follow-up data. As will be described in the 'Results' section, a further 17 patients were not included in the analysis due to insufficient or no DCIS in the sectioned tissue. The final number of patients was 51 (13 BCEs and 38 non-BCE). Demographic and treatment data are summarised for all patients ( $N = 135$ ) and the final group used for outcome analysis ( $N = 51$ ) in Supplementary Table 1.

Breast cancer screening was introduced into the UK in 1988, for women aged between 50 and 70 years, so many of the younger patients in this study presented symptomatically. Patients were subsequently treated off-protocol with lumpectomy (80%) or mastectomy (20%), adjuvant hormone therapy (56%), and adjuvant radiotherapy (74%) and 16% had both radiotherapy

and endocrine therapy (Tamoxifen). None of the patients received aromatase inhibitors.

After the review of H&Es, tissue microarrays were constructed from 135 patients, 2 mm dimension, with 40 cores per slide and 8 clinical slides total including 1 replicate, plus one control array with a mix of breast cancer and DCIS cores. Patients were primarily Caucasian, with age range of 34–75 (median age 55 years). The primary endpoint was any breast cancer event including ipsilateral or contralateral breast events and in rare cases metastases.

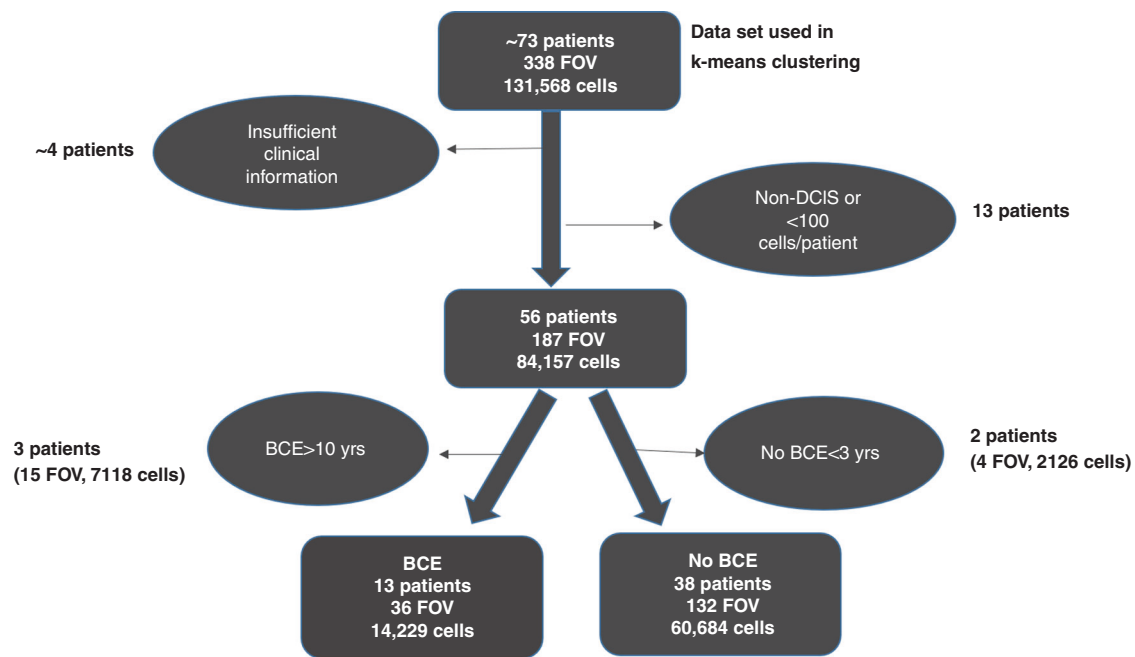
### Multiplexed immunofluorescence imaging of DCIS TMAs

Multiplexed immunofluorescence iterative staining of the DCIS TMAs was performed as previously described<sup>13,31</sup> using the Cell DIVE™ technology (Cytiva, Issaquah, WA). Briefly, slides were rehydrated, underwent a two-step antigen retrieval and were stained using a Leica Bond autostainer. After de-paraffinisation, and antigen retrieval,<sup>31</sup> the slides were incubated with antibodies at manufacturer recommended concentrations for 1 h at room temperature. All antibodies were validated per the protocol described previously.<sup>13,31</sup> Where possible, antibodies used in a clinical IHC lab were included in the screenings. After selection, each antibody was conjugated with either Cy3, Cy5 or Cy7 bis-NHS-ester dyes using standard protocols as previously described.<sup>31</sup> Supplementary Table 2 shows the antibodies, clones and conjugates used in the study.

All the sections underwent multiplexed staining for 32 markers (staining for one marker (S6) was repeated due to poor staining in the second round) using antibodies at concentrations as listed in Supplementary Table 3. Briefly the 32 markers and staining rounds were as follows: Round 1: CK5/6, Her4; Round 2: ABCG2, PTEN, S6; Round 3: CD20, S6 (repeated), CKAE1; Round 4: PR, ER, NaKATPase; Round 5: CK19, ALDH1, PCK26; Round 6: CD4, cMET; Round 7: CD44v6, HER2; Round 8: CDCP1, p53; Round 9: CK15, Cox2; Round 10: VEGFR2, ABCB1; Round 11: HTF9c, CD10; Round 12: MRP4, SLC7A5; Round 13: EGFR, p21; Round 14: MRP5, CEACAM5; Round 15 Ki67 (note that in total, 6 background imaging rounds were also included). Example stains and FOV for all markers are shown in Supplementary Fig. 1. All biomarkers stained as expected and were included in univariate analysis, however, staining for EGFR was very weak in some cases and was excluded from later analyses.

### Image processing, single cell segmentation

Cells in the epithelial and stromal compartments were then segmented using DAPI, pan-cytokeratin, S6, and NaKATPase, as previously described.<sup>13,31</sup> Several quality control steps were then conducted, including visual review and manual scoring of tissue quality and segmentation for every image. Images with poor quality staining or too few cells or lacking DCIS histology were excluded from data analysis. Also, additional cell filtering was conducted on the segmented images using the following criteria: (1) epithelial cells were required to have 1–2 number of nuclei; (2) each sub-cellular compartment (nucleus, membrane, cytoplasm) area had to have >10 pixels and <1500 pixels; and (3) cells in each round of staining had to have excellent alignment with the first round of staining (QC score = 1). For this, an automated QC score was generated for every cell in each imaging round by correlating baseline DAPI images with all corresponding DAPI images from other multiplexing rounds. A perfect score of 1 indicated perfect registration, no cell loss and no cell movement. A score of 0 indicated complete loss of that cell after baseline imaging. Supplementary Table 4 provides details of all the QC steps and the final number of cells that underwent statistical analysis. After the QC steps, the data underwent exposure time correction, log2 transformation to handle the skewness of the marker intensities, and normalisation to remove any slide batch effects. Additional patient level filtering was also applied as shown in Fig. 1 to ensure



**Fig. 1** A flow chart representing the total number of cells and fields of view (FOV) that were analysed and the filtering process that was used for breast cancer events (BCE) analysis. The end result of this process led to the analysis of 51 patients (13 patients with BCEs and 38 without BCEs).

a final dataset with complete clinical data. Median nuclear intensity was used for ER, PR, p21, Ki67 and median intensity of the whole cell for the remaining markers.

#### Statistical analysis

Correlation analysis was conducted for all biomarkers to determine co-expression patterns. For univariate and multivariate outcome analysis, the primary clinical endpoint was development of a BCE. Benjamini–Hochberg procedure with a false discovery rate 0.2 was applied to determine statistical significance of the association with BCE. For outcome analysis, images with insufficient DCIS content ( $\%DCIS/(\%DCIS + \%Normal) < 0.5$ ), and patients with less than 100 cells were excluded. Also, patients who were reported as having non-BCE, but had a follow-up time less than 3 years were excluded. Similarly, those who were reported as having BCE, but  $\geq 10$  years were also excluded to remove the possibility of a second primary.

In the first step of univariate analysis, violin plots of every biomarker distribution in the BCE and non-BCE groups were generated. Mean of the median cell intensities per marker were used for patient level aggregation and *t*-test was used to evaluate the mean differences between groups.

Unsupervised k-means clustering was applied with number of group  $k = 2, \dots, 15$  for all markers and subsets of markers. Extreme values (1% on both tails) were capped and standardised with zero mean and single standard deviation to remove unit effect of each marker. Consensus clustering<sup>32</sup> was used to determine the best number of distinct clusters. Consensus clustering involves repeating k-means clustering in a subset of the data and measuring how consistently the data separates into groups. PAC (proportion of ambiguously clustered)<sup>33</sup> and a visual check of the heatmap were used to evaluate the cluster separation. Following cluster analysis, the proportion of cells in each cluster, was determined for each patient (patient cluster profile). The patient cluster profile was evaluated in multivariate logistic regression models to determine probability of BCE for a given patient cluster profile. Leave-one-out cross validation was also performed to confirm the results. Log-rank test and Kaplan–Meier plot were generated to evaluate the effectiveness of the model.

#### RESULTS

Generation of a multi-step analyses workflow for the development of a DCIS breast cancer event score using multiplexed cell analysis data

As described in the ‘Methods’ section, the study cohort consisted of 13 patients who had a breast cancer event (BCE) within 10 years (median 2.5 years), and 38 patients who did not develop BCE within 3–10 years (median 9 years) (Supplementary Table 1). Survival analysis of the entire cohort compared to the study cohort showed no difference in outcome (Supplementary Fig. 2).

To develop a risk score for DCIS cases from multiplexed tissue, we developed a comprehensive multi-step data analyses and cell filtering workflow, as shown in Supplementary Table 3. Briefly, of the ~2 million cell-like objects (which included non-cell objects), which were segmented using the previously described membrane, cytoplasm and nuclear markers, 40% were classified as potential epithelial cells based on their location within cytokeratin positive tissue regions. After applying the filtering criteria described in the method section, 131,568 nucleated epithelial cells were used for clustering analysis. Of those, 74,913 cells with matching clinical data were used for outcome analysis.

Clinical features such as patient age and menopause status and histological features such as grade and comedonecrosis were not associated with breast cancer events (BCE). Further, no difference in treatment regimens were found between the non-BCE and BCE groups (Fig. 2) and thus treatment was not included in the final model.

Example images for all markers are shown in Supplementary Fig. 1. Univariate correlation analysis plots between all markers, irrespective of group, in Supplementary Fig. 3. To initially analyse the association between marker expression and BCE, violin plots were generated for each of the markers, shown in Supplementary Fig. 4. In univariate analysis, higher HER2 and lower ER and PR expression were associated with BCE (based on Benjamini–Hochberg procedure with a false discovery rate 0.2). EGFR was associated with BCE but was excluded from the analysis due to overall poor and inconsistent staining quality. The remainder of the markers that were not associated with BCE were as follows: ABCB1, ABCG2, ALDH1, CDCP1, CD10, CD44v6,

		N	Y	
	Surg only	17	6	23
	Surg + HT	10	4	14
	Surg + RT	4		4
	Surg + HT + RT	6	2	8
		37	12	49

**Fig. 2 Characteristics of the clinical cohort.** Incidence of breast cancer events (BCEs) in patients with hormone therapy (HT), radiotherapy (RT), or surgery (Surg) alone or in combination. N-No BCE, Y-Yes for BCE.

CEACAM5, CK-15, CK-19, CK-56, CK-AE1, CK-PCK26, cMET, HER4, HTF9C, MRP4, PTEN, MRP5, NaKATPase, p53, p21, S6, SLC7A5 and VEGFR2. Notably, Ki67 and COX2 were not associated with BCEs ( $P = 0.561$  and  $P = 0.851$ , respectively).

#### K-Means clustering of cellular protein expression

Based on the initial univariate results, we conducted cell cluster analysis with subsets of all epithelial biomarkers along with HER2 and ER to explore which biomarkers were co-expressed with this clinical phenotype and significantly correlated with outcome (PR was excluded as it was highly correlated with ER expression). Although neither SLC7A5 nor cMET were independently correlated with BCE in univariate analysis, when clustered along with ER and HER2 positive cells they were consistently aligned. Using consensus clustering of ER, HER2, SLC7A5 and cMET, 6 clusters with varying levels of each marker were found to provide the best separation.

Figure 3a shows the cluster distribution heatmap for all 6 clusters and a breakdown of cluster profiles for the BCE and non-BCE groups is shown in the adjacent table. As shown in Fig. 3b, cluster 2, characterised by higher expression of ER and lower HER2, SLC7A5 and cMET, accounted for 17.6% of cells in the non-BCE group and was strongly associated with lack of BCE ( $P = 0.001$ ), compared to 3% of cluster 2 cells in the BCE group. Similarly, cluster 4 (higher expression of ER and cMET, and lower HER2 and SLC7A5) accounted for 16.5% of cells in the non-BCE group and was also associated with lack of BCE ( $P = 0.034$ ), vs 6.3% in the BCE group. Cluster 6 with higher expression of HER2, lower SLC7A5, ER and cMET, accounted for 33% of cells in the BCE group and was associated with high risk of BCE ( $P = 0.018$ ), vs 11.4% of cells in the non-BCE group. Cluster 5, with both high HER2 and SLC7A5, low-moderate cMET and low ER accounted for 24.5% of cells in the BCE group (vs. 8.5% in the non-BCE group), associated with high risk of BCE ( $P = 0.072$ ), but suggesting a possible impact of SLC7A5 in determination of BCE. Example images including virtual H&E, biomarker stains and associated cluster plots for each cluster and combinations of clusters are shown in Fig. 4a, b. Cluster 1, which had low expression for all 4 markers or Cluster 3, which had high SLC7A5 and low HER2 were not significantly correlated with the outcome ( $P = 0.742$  and  $0.221$ , respectively).

Development of logistic regression-based algorithm for predicting risk of BCE using cell clusters (Escore)

In order to further assess the clinical utility of the cell clusters, logistic regression analysis was performed using combinations of clusters that were correlated with BCE (clusters 2 + 4 and clusters 5 + 6). The model gave an AUC of 0.79 (0.74 with leave-one-out cross validation). This analysis was further converted into an

expression score of the likelihood of BCE, 'Escore', based on the predicted probability of BCE from the logistic regression model (Fig. 5a).  $\text{Escore} = 1.77 * (\% \text{Clus5\&6}) - 2.78 * (\% \text{Clus2\&4}) > 13$  was the criteria for the high-risk BCE with sensitivity (TPR) of 77% and a specificity (TNR) of 79%. The threshold was based on the predicted probability of BCE being higher than the group BCE rate of 0.255 (13 divided by 51). Figure 5b shows the disease-free-interval analysis using Kaplan–Meier plots for the two groups. As shown in these plots, binary categorisation of the Escore results in clear separation of the survival curves ( $P = 5E-05$ ) with low scores being associated with marked decrease in likelihood of BCE. In initial validation using the leave-one-out cross validation, Escore remained significantly associated with BCE ( $P = 0.006$ ).

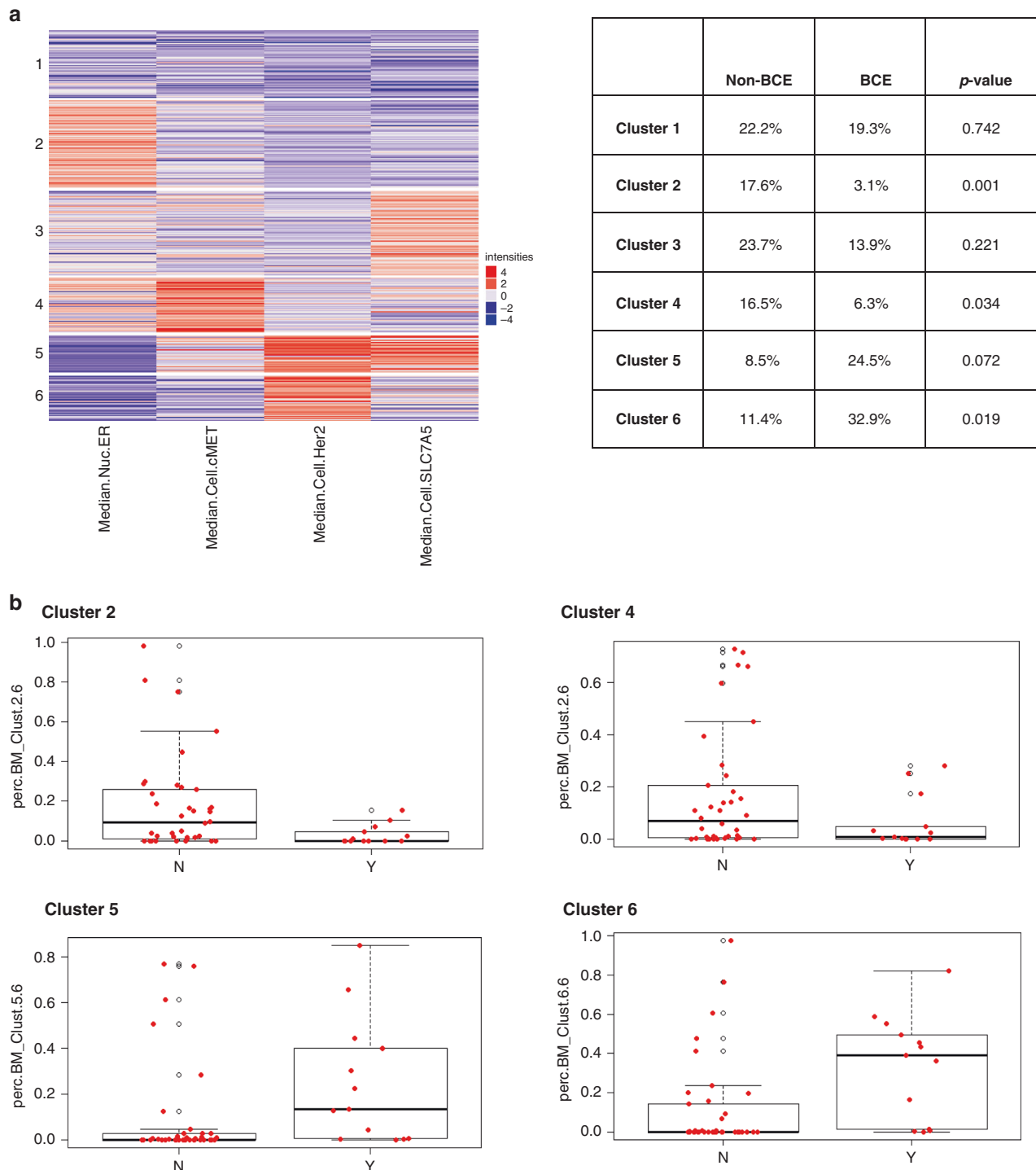
The cohort had 10 patients for whom data was available on duplicate cores. Analysis of these cores showed consistent results; none of the patients would have been re-classified based on these additional cores. This suggests that even in this small cohort and limited tissue sample availability, Escore was not affected by heterogeneity; however, the sample size is too small for a definitive conclusion. A recent database update resulted in the reclassification of BCE status of 4 patients from non-BCE to BCE. As these BCEs occurred 9.6–16.2 years after initial diagnosis, these might be potentially new diseases. Two of these 4 patients were in the low risk group, as determined by Escore.

## DISCUSSION

Biological heterogeneity in cancer is well recognised,<sup>12</sup> leading to the understanding that not all cancers need to be treated aggressively. In the case of invasive breast cancer, gene expression assay-based trials such as the MINDACT (microarray in the determination of adjuvant chemotherapy) and TAILORx (trial assigning Individualised options for treatment Rx) have documented that a significant number of women can safely avoid chemotherapy.<sup>34,35</sup> Both assays were good at identifying classes of patients that benefit from chemotherapy (high clinical and high molecular risk groups) and that can safely avoid chemotherapy (low clinical and low molecular risk groups).

Epidemiological studies have documented that overall survival rates for DCIS are around 95% at 10 years.<sup>36</sup> Therefore, it is natural to seek to identify categories of patients for whom therapy can be reduced. Additionally, there is great concern about 'overdiagnosis' and hence overtreatment screen detected DCIS.<sup>37</sup> DCIS has been traditionally treated with surgery followed by hormonal therapy and or radiotherapy to the breast to prevent recurrence of DCIS or development of invasive cancer. The current clinical trials (LORIS, LORD and COMET) are enrolling patients on histological features<sup>7,8,10</sup>; this in part due to lack of good molecular markers. One of the major limitations of the immunohistochemical or mRNA panels is the amount of tissue required for analysis. This is particularly true in cases where important management decisions are going to be made on tiny fragment of "tumour" tissue in needle core biopsies. In an effort to minimise the tissue requirements, we have used multiplex immunofluorescence (MxIF, Cell Dive™) to identify cellular parameters associated with BCE.

The current study is based on analysis of a single section of the tissue microarray (TMA) from patients with DCIS. This is a potential limitation of the study in so far as DCIS is a heterogeneous disease and a TMA core does not adequately represent the extent of disease. This may also explain why data from at least 10 patients could not be used in the final analysis due to absence of DCIS lesions in the section. That being said, this is an exploratory study on cell protein phenotypes and TMAs provide a convenient approach to analyse a larger number of patients. Thirty-two markers were analysed on a single paraffin section using 15 cycling rounds of staining and imaging. This is a major strength of the study. However, the analysis also resulted in data loss due to the requirement that field be composed almost entirely of DCIS



**Fig. 3 Contribution of cell clusters to BCE and non-BCE.** **a** Heatmap depicting the results of the unsupervised k-means clustering. The use of four markers (HER2, SLC7A5, ER, and cMET) resulted in the identification of six clusters (1–6). Cluster 2 cells were enriched for ER positivity, and cluster 4 shows ER and cMET positivity. HER2 expression was observed in both clusters 5 and 6. The likelihood of breast cancer events (BCE) associated with each of the clusters is also shown. **b** Box plots showing the association of the clusters with likelihood of BCE. BCE patients tended to have very little cluster 2 and 4 type cells and more of cluster 5 and 6 type cells. After k-means clustering is performed, the proportion of each cell type (clusters) in each patient is calculated. Comparison of cluster profiles by BCE showed that cluster 2 and 4 cells appeared more in non-BCE patients while cluster 5 and 6 appeared more in BCE patients.

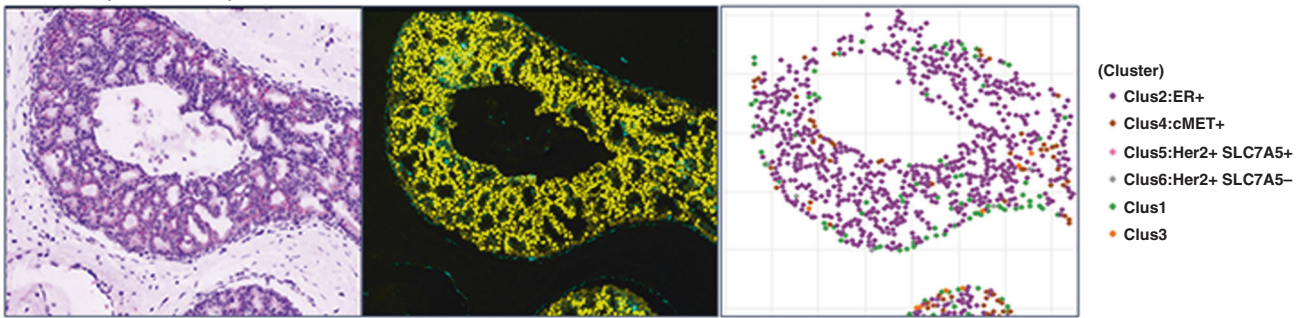
cells. This criterion was used to reduce the impact of normal (contaminating) epithelial elements and made it easier to analyse the data as it did not require cell-level classification of the lesions. Further improvement of the analysis algorithms, such as machine learning based approaches for sub-regions of interest, would likely

reduce such losses. Further validation studies to confirm the accuracy and reliability of this prognostic signature are necessary and are being planned.

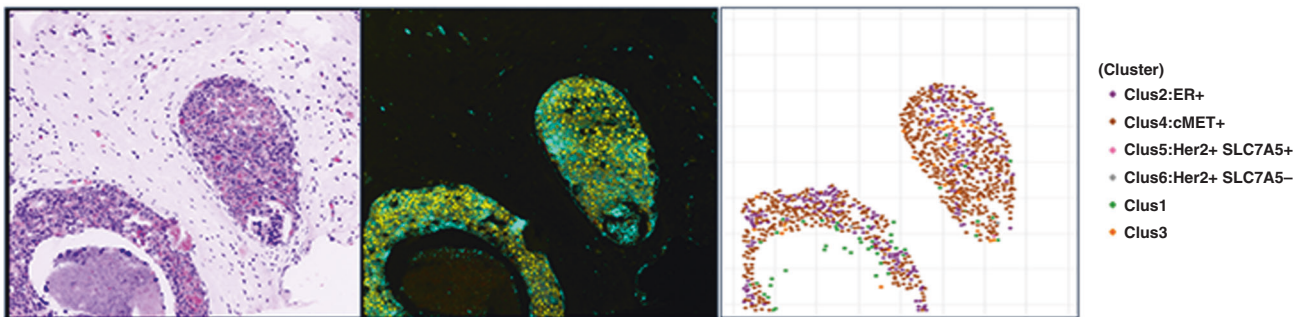
The association of grade with recurrence was non-significant. However, it must be noted that grade was not significant in the



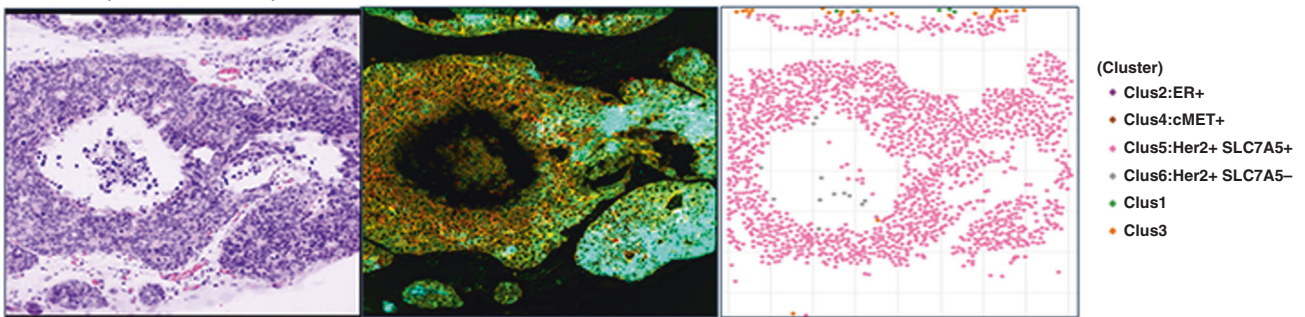
**Cluster 2 (ER+ cMET-) dominant**



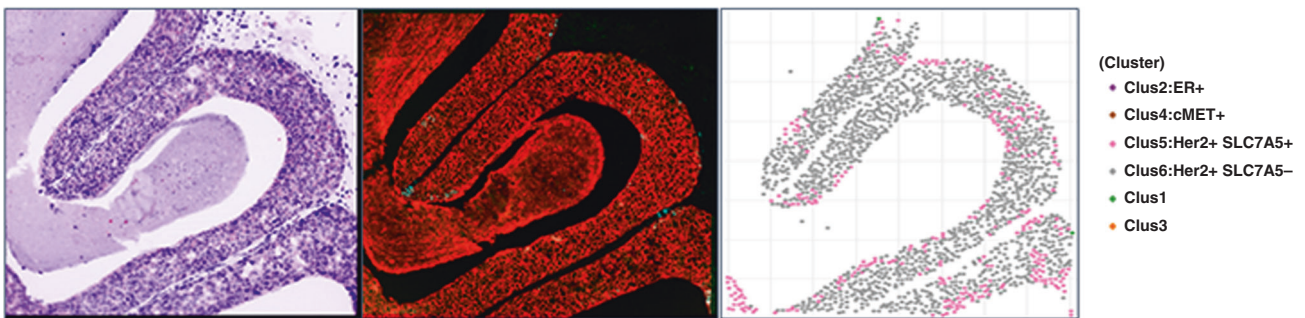
**Cluster 4 (ER+ cMET+)**



**Cluster 5 (Her2+ SLC7A5+)**



**Cluster 6 (Her2+ SLC7A5-) dominant**

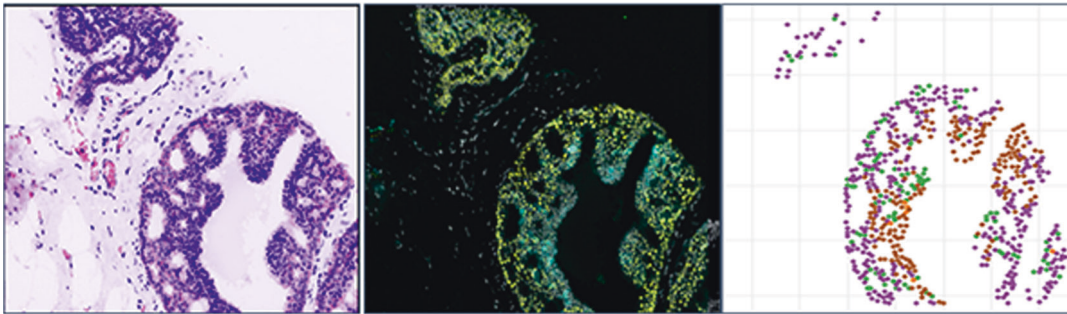


**Fig. 4** Continued.

analysis of 108,196 patients from the SEER database in multi-variate analysis.<sup>36</sup> In univariate analysis, only ER, PR, and HER2 were associated with likelihood of BCE. This is consistent with prior literature and suggests that the result observed herein can be potentially generalisable. None of the other markers analysed, including grade were associated with BCE. This may be due to small sample size, and insufficient power to detect additional prognostic role of features with a weaker influence on outcome.

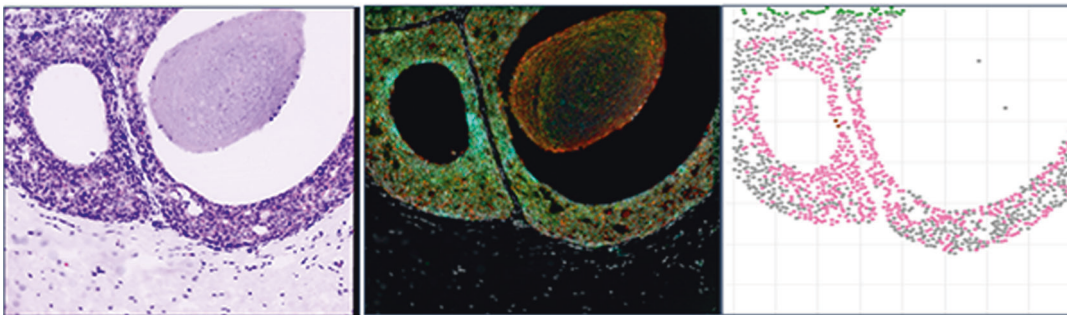
The expression of both Ki67 and COX2, which in the context of DCIS has been associated with proliferation,<sup>9</sup> was not associated with BCE. These results are in contrast to prior studies, where Ki67 and COX2 expression were associated with recurrence.<sup>18</sup> Importantly, mRNAs of the proliferation related genes play an important role in Oncotype Dx DCIS score.<sup>19</sup> One possible explanation of these differences is that we used an automated quantitation method of signal intensity for all cells included in the analysis and

Cluster 2 and cluster 4



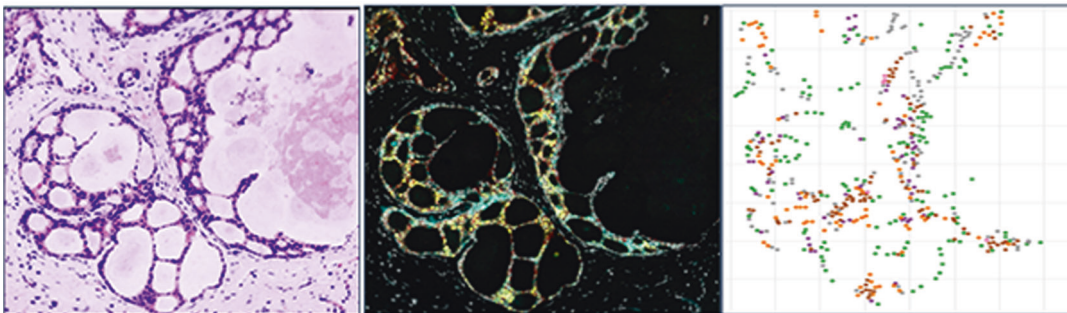
- (Cluster)
- Clus2:ER+
  - Clus4:cMET+
  - Clus5:Her2+ SLC7A5+
  - Clus6:Her2+ SLC7A5-
  - Clus1
  - Clus3

Cluster 5 and cluster 6



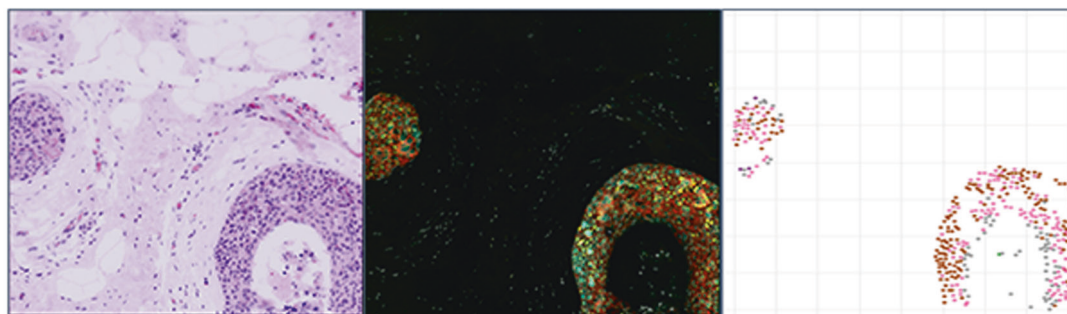
- (Cluster)
- Clus2:ER+
  - Clus4:cMET+
  - Clus5:Her2+ SLC7A5+
  - Clus6:Her2+ SLC7A5-
  - Clus1
  - Clus3

All clusters



- (Cluster)
- Clus2:ER+
  - Clus4:cMET+
  - Clus5:Her2+ SLC7A5+
  - Clus6:Her2+ SLC7A5-
  - Clus1
  - Clus3

Cluster 4 and cluster 5



- (Cluster)
- Clus2:ER+
  - Clus4:cMET+
  - Clus5:Her2+ SLC7A5+
  - Clus6:Her2+ SLC7A5-
  - Clus1
  - Clus3

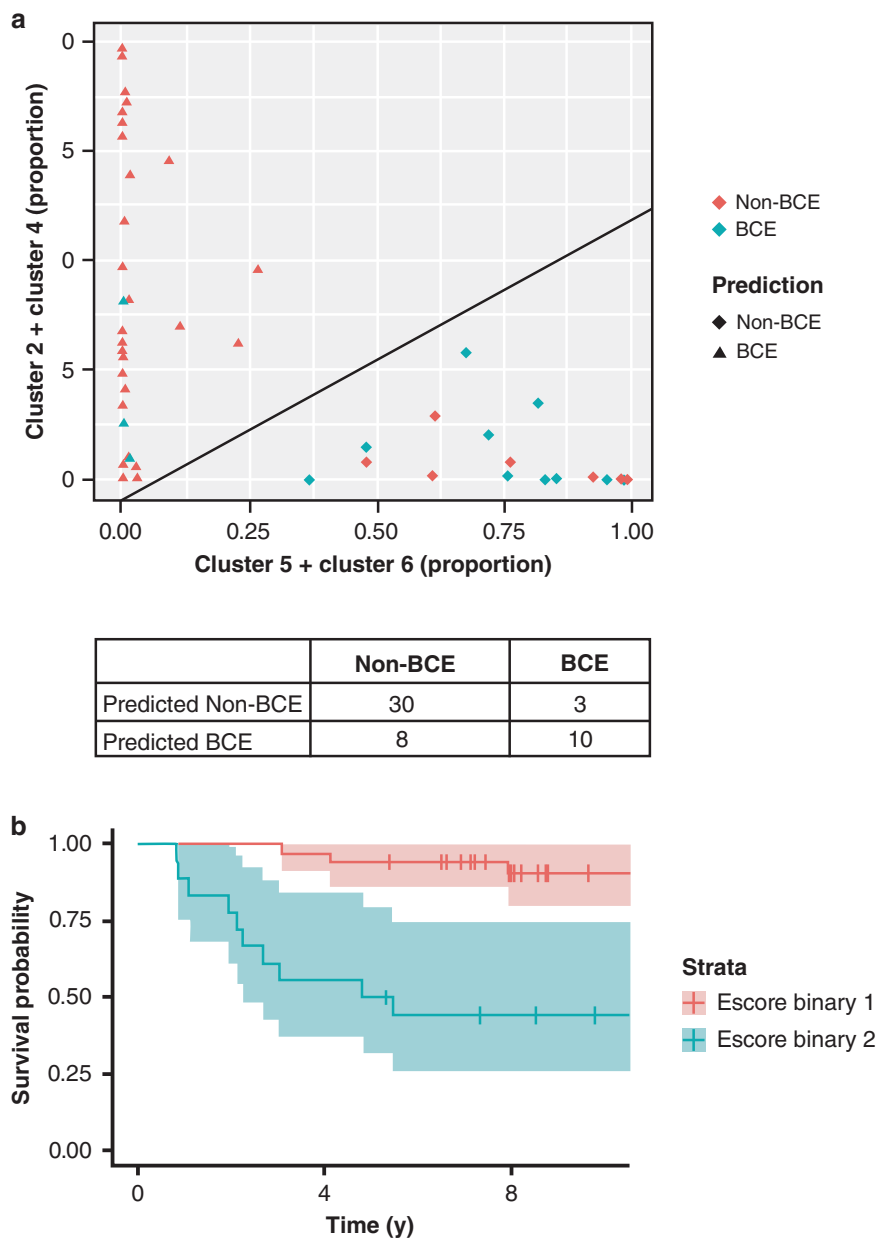
**Fig. 4 Virtual H&E images for each example cluster, biomarker stains and cluster plot of the cells expressing those markers at different levels of intensity.** Ducts can either show dominant expression of a single cluster or mixed expression pattern of clusters.

our final unit of measurement for each cell was median nuclear intensity. While all levels of nuclear intensity are considered positive in the assessment of Ki67,<sup>38,39</sup> we could not apply a pathologist determined cut-off and unsupervised analysis of biomarker intensities versus outcome was conducted. Furthermore, Ki67 levels in DCIS are low and are not

'thresholded' in the DCIS score, in contrast to the 21-gene recurrence score.<sup>19</sup>

Although the expression of SLC7A5 and cMET was not significant in univariate analysis, cluster 4 with high expression of cMET, moderate/high ER, low HER2 and SLC7A5 was associated with low risk of BCE. However, cluster 5 and 6, with high HER2





**Fig. 5 Model performance for prediction of BCE and non-BCE. a** Classification model was developed using logistic regression to predict BCE and non-BCE using two input variables, one comprising of cluster 2 and 4, and the other comprising of cluster 5 and 6. The model had a sensitivity of 77%, specificity of 79% with an error rate of 21.6% and AUC of 0.785. The AUC in the leave one out cross-validation was 0.739. **b** Escore was developed from the classification model and binary Escore was used for log-rank analysis and data represented using Kaplan-Meier plots.

expression resulted in increased risk of BCE. Of note, cMET and SLC7A5 have not been previously implicated in prognostication of DCIS. In invasive cancer, cMET overexpression is seen in metastatic tumours and the amino acid transporter SLC7A5 is a component of the MammaStrat™ signature and more recently shown to be a key therapeutic target in ER+ breast cancer.<sup>40</sup> These data support the relevance of cMET and SLC7A5 proteins in the biology of breast cancer.

The combination of cluster-based expression scores of ER/HER2/cMET and SLC7A5 markers contributed to development of the Escore algorithm, which was significantly predictive of likelihood of BCE ( $P = 0.00005$ ). In preliminary validation using leave one out cross-validation (LOOCV) method, Escore remained significant ( $P = 0.006$ ). In addition, analysis of duplicate cores from the same patients resulted in similar Escore.

There are a number of limitations to the current study. The analysis was based on retrospective cohort with negative margins (margins status could not be reconfirmed). The analysis was also based on ‘median cell intensity’ and ‘median nuclear intensity’ and the cut-offs were not optimised in the current study. This could have possibly caused lack of significant association of some biomarkers, such as Ki67 and Cox2 with BCE. Lastly, the study is exploratory in nature and the findings need further validation in additional cohorts to understand the value of spatial multiplex immunofluorescence analysis and utility of the Escore. However, we believe that the analysis at single cell level provides considerable strengths and the development of Escore is based on the expression of biomarkers known to play important role in breast cancer. It is important to note that the cohort used herein was matched for treatment. There was no difference in treatment



regimens between the BCE and non-BCE patients, who were split approximately 50/50 for treatment and non-treatment regimens (Fig. 2). Further analyses will include replication of the algorithm using Cell Dive™ that necessitates use of multiple markers for cell segmentation as well simpler methods using just the four markers and cell-based analysis. Success in generating the EScore using simple(r) IHC methods could result in rapid dissemination of the results and their implementation in clinical practice.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions of Alex Corwin for Cell DIVE imaging workflows and quality control and Sean Dinn and Eric Williams for antibody conjugations.

## AUTHOR CONTRIBUTIONS

Concept, design, interpretation and execution of the entire project: S.S.B., Y.G.-P., M.G., P.H.T., A.H. and F.G. Statistical and bioinformatics analyses and interpretation: S.C., M.T., M.Z. and Y.S. Manuscript writing: S.S.B., S.C., Y.G.-P., P.H.T., M.G., A.H. and F.G. All authors approved the manuscript.

## ADDITIONAL INFORMATION

**Ethics approval and consent to participate** The study was performed in accordance with the Declaration of Helsinki. Waiver of IRB was obtained from Indiana, and Oxford Universities.

**Consent to publish** Not applicable.

**Data availability** Additional information can be found in Supplemental data section. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests** A patent disclosure has been filed at IU for the EScore. The authors (S.S.B., Y.G.P., F.G., S.C. and A.L.H.) have been listed as inventors. S.C., Y.S., C.C., E.M., A.S., M.Z., M.G. and F.G. are employees of GE. The remaining authors do not have any competing interests.

**Funding information** This study is supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA194600 to S. Badve, M. Gerdes and F. Ginty. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41416-020-01216-6>.

**Note** This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution 4.0 International (CC BY 4.0).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Ebctcg, McGale, P., Taylor, C., Correa, C., Cutter, D., Duane, F. et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet* **383**, 2127–2135 (2014).
2. Czick, J., Sestak, I., Pinder, S. E., Ellis, I. O., Forsyth, S., Bundred, N. J. et al. Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: long-term results from the UK/ANZ DCIS trial. *Lancet Oncol.* **12**, 21–29 (2011).
3. Wapnir, I. L., Dignam, J. J., Fisher, B., Mamounas, E. P., Anderson, S. J., Julian, T. B. et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J. Natl. Cancer Inst.* **103**, 478–488 (2011).
4. Sanders, M. E., Schuyler, P. A., Dupont, W. D. & Page, D. L. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer* **103**, 2481–2484 (2005).

5. Jones, J. L. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. *Breast Cancer Res.* **8**, 204 (2006).
6. Collins, L. C., Tamimi, R. M., Baer, H. J., Connolly, J. L., Colditz, G. A. & Schnitt, S. J. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer* **103**, 1778–1784 (2005).
7. Elshof, L. E., Tryfonidis, K., Slaets, L., van Leeuwen-Stok, A. E., Skinner, V. P., Dif, N. et al. Feasibility of a prospective, randomised, open-label, international multi-centre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ—The LORD study. *Eur. J. Cancer* **51**, 1497–1510 (2015).
8. Fallowfield, L., Francis, A., Catt, S., Mackenzie, M. & Jenkins, V. Time for a low-risk DCIS trial: harnessing public and patient involvement. *Lancet Oncol.* **13**, 1183–1185 (2012).
9. Groen, E. J., Elshof, L. E., Visser, L. L., Rutgers, E. J. T., Winter-Warnars, H. A. O., Lips, E. H. et al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). *Breast* **31**, 274–283 (2017).
10. Grimm, L. J., Ryser, M. D., Partridge, A. H., Thompson, A. M., Thomas, J. S., Wesseling, J. et al. Surgical Upstaging rates for vacuum assisted biopsy proven DCIS: implications for active surveillance trials. *Ann. Surg. Oncol.* **24**, 3534–3540 (2017).
11. Badve, S., A'Hern, R. P., Ward, A. M., Millis, R. R., Pinder, S. E., Ellis, I. O. et al. Prediction of local recurrence of ductal carcinoma in situ of the breast using five histological classifications: a comparative study with long follow-up. *Hum. Pathol.* **29**, 915–923 (1998).
12. Badve, S. & Gokmen-Polar, Y. Tumor heterogeneity in breast cancer. *Adv. Anat. Pathol.* **22**, 294–302 (2015).
13. Gerdes, M. J., Gokmen-Polar, Y., Sui, Y., Pang, A. S., LaPlante, N., Harris, A. L. et al. Single-cell heterogeneity in ductal carcinoma in situ of breast. *Mod. Pathol.* **31**, 406–417 (2018).
14. Harrison, B. T., Hwang, S., Partridge, A., Thompson, A. & Schnitt, S. J. Variability in diagnostic threshold for comedo necrosis among pathologists: implications for patient eligibility for active surveillance trials of DCIS. *Mod. Pathol.* **31**, 70 (2018).
15. Solin, L. J., Gray, R., Hughes, L. L., Wood, W. C., Lowen, M. A., Badve, S. S. et al. Surgical excision without radiation for ductal carcinoma in situ of the breast: 12-year results from the ECOG-ACRIN E5194 study. *J. Clin. Oncol.* **33**, 3938–3944 (2015).
16. Sagara, Y., Mallory, M. A., Wong, S., Aydogan, F., DeSantis, S., Barry, W. T. et al. Survival benefit of breast surgery for low-grade ductal carcinoma in situ: a population-based cohort study. *JAMA Surg.* **150**, 739–745 (2015).
17. Kerlikovske, K., Molinaro, A. M., Gauthier, M. L., Berman, H. K., Waldman, F., Bennington, J. et al. Biomarker expression and risk of subsequent tumors after initial ductal carcinoma in situ diagnosis. *J. Natl. Cancer Inst.* **102**, 627–637 (2010).
18. Bremer, T., Whitworth, P. W., Patel, R., Savala, J., Barry, T., Lyle, S. et al. A biological signature for breast ductal carcinoma in situ to predict radiotherapy benefit and assess recurrence risk. *Clin. Cancer Res.* **24**, 5895–5901 (2018).
19. Solin, L. J., Gray, R., Baehner, F. L., Butler, S. M., Hughes, L. L., Yoshizawa, C. et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *J. Natl. Cancer Inst.* **105**, 701–710 (2013).
20. Rakovitch, E., Nofech-Mozes, S., Hanna, W., Sutradhar, R., Baehner, F. L., Miller, D. P. et al. Multigene expression assay and benefit of radiotherapy after breast conservation in ductal carcinoma in situ. *J. Natl. Cancer Inst.* **109**, djw256 (2017).
21. Gokmen-Polar, Y., Nakshatri, H. & Badve, S. Biomarkers for breast cancer stem cells: the challenges ahead. *Biomark. Med.* **5**, 661–671 (2011).
22. Wright, H. J., Hou, J., Xu, B., Cortez, M., Potma, E. O., Tromberg, B. J. et al. CDCP1 drives triple-negative breast cancer metastasis through reduction of lipid-droplet abundance and stimulation of fatty acid oxidation. *Proc. Natl. Acad. Sci. USA* **114**, E6556–E6565 (2017).
23. Yang, C., He, P., Liu, Y., He, Y., Yang, C., Du, Y. et al. Down-regulation of CEACAM1 in breast cancer. *Acta Biochim. Biophys. Sin. (Shanghai)*. **47**, 788–794 (2015).
24. Wakasugi, E., Kobayashi, T., Tamaki, Y., Ito, Y., Miyashiro, I., Komoike, Y. et al. p21 (Waf1/Cip1) and p53 protein expression in breast cancer. *Am. J. Clin. Pathol.* **107**, 684–691 (1997).
25. Nasir, A., Holzer, T. R., Chen, M., Man, M. Z. & Schade, A. E. Differential expression of VEGFR2 protein in HER2 positive primary human breast cancer: potential relevance to anti-angiogenic therapies. *Cancer Cell Int.* **17**, 56 (2017).
26. Hicks, D. G., Janarthanan, B. R., Vardarajan, R., Kulkarni, S. A., Khoury, T., Dim, D. et al. The expression of TRMT2A, a novel cell cycle regulated protein, identifies a subset of breast cancer patients with HER2 over-expression that are at an increased risk of recurrence. *BMC Cancer* **10**, 108 (2010).
27. Bartlett, J. M., Thomas, J., Ross, D. T., Seitz, R. S., Ring, B. Z., Beck, R. A. et al. Mammostrat as a tool to stratify breast cancer patients at risk of recurrence during endocrine therapy. *Breast Cancer Res.* **12**, R47 (2010).
28. Kochel, T. J., Reader, J. C., Ma, X., Kundu, N. & Fulton, A. M. Multiple drug resistance-associated protein (MRP4) exports prostaglandin E2 (PGE2) and

- contributes to metastasis in basal/triple negative breast cancer. *Oncotarget* **8**, 6540–6554 (2017).
29. Mao, Q. & Unadkat, J. D. Role of the breast cancer resistance protein (BCRP/ABCG2) in drug transport—an update. *AAPS J.* **17**, 65–82 (2015).
  30. Delou, J. M. A., Vignal, G. M., Indio-do-Brasil, V., Accioly, M. T. S., da Silva, T. S. L., Piranda, D. N. et al. Loss of constitutive ABCB1 expression in breast cancer associated with worse prognosis. *Breast Cancer (Dove Med Press)*. **9**, 415–428 (2017).
  31. Gerdes, M. J., Sevinsky, C. J., Sood, A., Adak, S., Bello, M. O., Bordwell, A. et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl Acad. Sci. USA*. **110**, 11982–11987 (2013).
  32. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
  33. Şenbabaoğlu, Y., Michailidis, G. & Li, J. Z. Critical limitations of consensus clustering in class discovery. *Sci. Rep.* **4**, 6207 (2014).
  34. Cardoso, F., van't Veer, L. J., Bogaerts, J., Slaets, L., Viale, G., Delaloge, S. et al. 70-Gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* **375**, 717–729 (2016).
  35. Sparano, J. A., Gray, R. J., Makower, D. F., Pritchard, K. I., Albain, K. S., Hayes, D. F. et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N. Engl. J. Med.* **379**, 111–121 (2018).
  36. Narod, S. A., Iqbal, J., Giannakeas, V., Sopik, V. & Sun, P. Breast cancer mortality after a diagnosis of ductal carcinoma in situ. *JAMA Oncol.* **1**, 888–896 (2015).
  37. Esserman, L. J., Thompson, I. M., Reid, B., Nelson, P., Ransohoff, D. F., Welch, H. G. et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol.* **15**, e234–e242 (2014).
  38. Leung, S. C. Y., Nielsen, T. O., Zabaglo, L., Arun, I., Badve, S. S., Bane, A. L. et al. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. *NPJ Breast Cancer* **2**, 16014 (2016).
  39. Leung, S. C. Y., Nielsen, T. O., Zabaglo, L. A., Arun, I., Badve, S. S., Bane, A. L. et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. *Histopathology* **75**, 225–235 (2019).
  40. El Ansari, R., Craze, M. L., Miligy, I., Diez-Rodriguez, M., Nolan, C. C., Ellis, I. O. et al. The amino acid transporter SLC7A5 confers a poor prognosis in the highly proliferative breast cancer subtypes and is a key therapeutic target in luminal B tumours. *Breast Cancer Res.* **20**, 21 (2018).