Single-cell heterogeneity in ductal carcinoma *in situ* of breast

Michael J Gerdes¹, Yesim Gökmen-Polar², Yunxia Sui¹, Alberto Santamaria Pang¹, Nicole LaPlante⁵, Adrian L Harris³, Puay-Hoon Tan⁴, Fiona Ginty¹ and Sunil S Badve²

¹GE Global Research, Diagnostics, Imaging and Biotechnology (DIBT), Niskayuna, NY, USA; ²Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN, USA;

³Department of Oncology, Cancer and Haematology Centre, Oxford University, Oxford, UK and ⁴Division of Pathology, Singapore General Hospital, Singapore

Heterogeneous patterns of mutations and RNA expression have been well documented in invasive cancers. However, technological challenges have limited the ability to study heterogeneity of protein expression. This is particularly true for pre-invasive lesions such as ductal carcinoma *in situ* of the breast. Cell-level heterogeneity in ductal carcinoma *in situ* was analyzed in a single 5 μ m tissue section using a multiplexed immunofluorescence analysis of 11 disease-related markers (EGFR, HER2, HER4, S6, pmTOR, CD44v6, SLC7A5 and CD10, CD4, CD8 and CD20, plus pan-cytokeratin, pan-cadherin, DAPI, and Na+K+ATPase for cell segmentation). Expression was quantified at cell level using a single-cell segmentation algorithm. K-means clustering was used to determine co-expression patterns of epithelial cell markers and immune markers. We document for the first time the presence of epithelial cell heterogeneity within ducts, between ducts and between patients with ductal carcinoma *in situ*. There was moderate heterogeneity in a distribution of eight clusters within each duct (average Shannon index 0.76; range 0–1.61). Furthermore, within each patient, the average Shannon index across all ducts ranged from 0.33 to 1.02 (s.d. 0.09–0.38). As the distribution of clusters within ducts was uneven, the analysis of eight ducts might be sufficient to represent all the clusters ie within- and between-duct heterogeneity. The pattern of epithelial cell clustering was associated with the presence and type of immune infiltrates, indicating a complex interaction between the epithelial tumor and immune system for each patient. This analysis also provides the first evidence that simultaneous analysis of both the epithelial and immune/stromal components might be necessary to understand the complex milieu in ductal carcinoma *in situ* lesions.

Modern Pathology (2018) 31, 406–417; doi:10.1038/modpathol.2017.143; published online 17 November 2017

Cellular heterogeneity presents a major problem in treating cancer. Tumor heterogeneity has been elucidated in detail using next-generation sequencing.^{1–4} Studies have identified substantial tumor heterogeneity with mutations in spatially distinct parts of the tumor in several cancer types. More recently, next-generation sequencing techniques using single tumor cells from invasive cancers have increased our ability to understand cellular heterogeneity.⁵ Understanding heterogeneity in protein expression is still a major challenge. This is even

more problematic in small pre-invasive lesions such as ductal carcinoma *in situ*.

Ductal carcinoma *in situ*, the most common type of pre-invasive lesion, is being detected with increasing frequency with the advent of mammographic screening. Surgery is the mainstay for the treatment of ductal carcinoma *in situ* and on the clinicopathological features, this may be followed by radiotherapy and/or endocrine therapy.⁶ The qualitative assessment of histological grade, expression of single protein biomarkers and more recently, mRNA analysis (Oncotype DX Breast DCIS Score)^{7–9} have been used to make these decisions. However, these approaches do not fully reflect the degree of cellular heterogeneity of ductal carcinoma *in situ*, cellular interactions with surrounding microenvironment and other cell types. Using quantitative multiplexed immunofluorescence of a single formalin-fixed paraffin-embedded tissue section,¹⁰ we have for the first time investigated the degree of cellular heterogeneity

Correspondence: Dr MJ Gerdes or Professor SS Badve, Department of Pathology and Laboratory Medicine, Indiana University School of Medicine, 350 West 11th Street, Indianapolis, IN 46202, USA. E-mail: gerdes@research.ge.com or sbadve@iupui.edu

⁵Current address: Clariant-NeoGenomics Laboratories, 31 Columbia, Aliso Viejo, CA 92656, USA.

Received 2 January 2017; revised 6 September 2017; accepted 11 September 2017; published online 17 November 2017

by analyzing the expression and co-expression of biologically relevant proteins at a single-cell and subcellular levels in tissues in ductal carcinoma *in situ* and correlations with immune response. In total, 14 protein markers were multiplexed on single formalin-fixed paraffin-embedded section per patient, including EGFR, HER2, HER4, pmTOR, CD44v6, SLC7A5, CD10, CD4, CD8 and CD20. Pan-cytokeratin, pan-cadherin, DAPI, Na+K+ATPase were used for epithelial and stromal cell segmentation. Markers were selected based on previously reported expression in ductal carcinoma *in situ* lesions and association with recurrence and included members of the ErbB family (EGFR, HER2, HER4);^{11–17} pmTOR, which has been associated with autophagy pathway in ductal carcinoma *in situ* and breast cancer.¹⁸ CD44v6 expression has been reported in expressed ductal carcinoma *in situ*, invasive cancers, but not in benign tissue¹⁹ and is also reported to be a putative stem cell marker.²⁰ It may also facilitate receptor tyrosine kinase signaling via binding to its heparin sulfate side chain.²¹ SLC7A5/LAT1 is an amino-acid exchange transporter that has not been studied in ductal carcinoma *in situ*, but is upregulated in breast cancers at both primary and metastatic sites.²² Increased expression correlates with poor prognosis and disruption leads to the inhibition of mTOR.²³ CD10 is expressed on myoepithelial cells and loss of expression may be associated with invasive disease.²⁴ Lymphocyte infiltration in triple-negative and HER2-positive subtypes have been associated with prognosis in ductal carcinoma *in situ* and breast cancer, with CD4+ T-helper cells responsible for facilitating antigen presentation; FOXP3+CD4+ regulatory T cells supporting proliferation of B cells and CD8+ cytotoxic T cells required for tumor destruction.^{25–28} The expression pattern of individual markers (both epithelial and stromal) in the current study conformed to that described in prior publications and hence is not described in detail herein.

Materials and methods

Antibody Screening and Selection

Antibody validation processes have been previously described.¹⁰ Briefly, for each protein target examined in the study, a minimum of three commercial initial antibodies were compared in parallel using Cy3 or Cy5 conjugated secondary antibodies (Jackson Immunologicals). Antibodies were assessed based on staining specificity and sensitivity characteristics using a multi-tissue array that contained breast and other tissue types (Pantomics #MNT241) and compared with patterns reported in literature and or the Human Protein Atlas (<http://www.proteinatlas.org/>). Where possible, antibodies used in clinical immunohistochemistry lab were included in the screenings. Slides were cleared and processed for antigen

retrieval as previously reported¹⁰ and all antibodies used at manufacturer's recommended concentrations with 1-h room temperature incubation. Specificity was determined based on localization pattern and comparison with nonspecific antibodies of the same isotype. The strongest staining and most specific antibodies were further tested to ensure that the antigen was not altered by the dye inactivation process by comparing staining on samples that were untreated or treated 1, 5 or 10 times with the dye inactivation process (used to multiplex) and subsequent washing in PBS before antibody staining. The 14 markers were: pan-cytokeratin, pan-cadherin, NaKATPase, SLC7A5, CD10, EGFR, HER2, HER4, S6, phospho-mTOR, CD44v6, CD4, CD8 and CD20.

Antibody Conjugation

After selection, each antibody was conjugated with either Cy3 or Cy5 bis-NHS-ester dyes using standard protocols as previously described.¹⁰ Where needed, antibodies were purified from BSA, glycerol or other stabilizing agents by protein A/G purification. Antibodies were initially conjugated at two different loading concentrations (2 and 4 dyes/antibody) and staining performance was compared with the secondary-detected antibodies used in the screening stage.

Clinical Cohort

Thirteen de-identified cases of ductal carcinoma *in situ* were selected from the archives at Singapore General Hospital. The criteria for the selection were as follows (1) they were excision specimens; (2) patients did not have invasive cancer; (3) patients had additional blocks of ductal carcinoma *in situ*; (4) multiple foci of ductal carcinoma *in situ* were present in the sections examined. Clinical information on the samples included were age at diagnosis, size (mm) of ductal carcinoma *in situ*. An initial histopathological review was performed to confirm the diagnosis of ductal carcinoma *in situ*. The diagnosis of ductal carcinoma *in situ* was confirmed independently by two pathologists. A second tier of histological analysis was performed using pseudo-H&E digital images of the stained slides to study the relationships between the epithelial and stromal components and to quantify the different elements (see below) within any given field. Multiple fields were quantified from each case. Data for the expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor (HER2) was obtained from the files of Singapore General Hospital. These stains were performed as a part of standard of care using ASCO-CAP guidelines to define positivity.^{29,30}

Multiplexed Immunofluorescence Staining

Multiplexed immunofluorescence staining was performed as previously described.¹⁰ Slides were

initially baked at 60 °C for 1 h before clearing with xylene and rehydration through a graded series of ethanol. Two-step antigen retrieval was performed as previously described¹⁰ and slides were blocked overnight in 4% BSA, 10% donkey serum, in PBS at 4 °C. After blocking, slides were stained with DAPI (1 µg/ml), cover-slipped with anti-fade mounting media, and background images collected in DAPI, eGFP, Cy3 and Cy5 filter channels. For initial imaging, a 10X objective was used, and the whole tissue imaged, followed by image stitching to create a composite image of the sample. The large representation of the whole tissue was then converted to a pseudo-H&E image and used for mapping specific fields for further interrogation for each slide. A minimum of 30 fields were chosen for each histological section and included ductal carcinoma *in situ*, and surrounding stromal areas. After selection, microscope coordinates were determined and exported for use during imaging calibrated to a 20X objective. Slides were then sequentially stained, imaged with a 20X objective, dyes inactivated and re-stained iteratively for a total of 14 markers under analysis (Supplementary Table S1). All dye-conjugated antibody stains were for 1 h at room temperature, and slides were processed for staining and washing using a Biocare autostainer. A summary table of the specific staining sequence and exposure times are provided in Supplemental data. All fields of view acquired were subsequently re-evaluated by a pathologist for % tumor, % normal epithelial, % immune along with pathological grading. The % tumor pertains to the % of any given field of view that is represented by ductal carcinoma *in situ*. Similarly, the amount in each field of view was quantified for normal epithelium and immune cells. The complete image set was then reviewed for tissue quality (tissue loss or damage) and image analysis segmentation output.

Image Processing and Statistical Analyses

Data preprocessing. After data collection, images from all rounds are registered to the initial round using the DAPI channel to determine any offsets. Images are then processed for field flattening and autofluorescence is subtracted. Several quality control steps were applied to ensure usable cell-level intensities. In addition, manual scoring of tissue quality and segmentation for every 'Field Of View' was performed; this resulted in exclusion of 35 field of views. For total 13 slides, 416 field of views were remaining for further analysis.

Single-cell analysis algorithm segments cells in the epithelial compartments using DAPI, pan-cytokeratin, S6, and NaKATPase, and just DAPI in the stromal compartment, as previously described.¹⁰ Expression of SLC7A5, CD10, EGFR, HER2, HER4, S6, phospho-mTOR, CD44v6 in the epithelial masked area was quantified at single-cell level and

metrics included mean intensity, maximum intensity and s.d. in the membrane, cytoplasm and nucleus. Quality checks and filtering at single-cell level were also performed. For example, epithelial cells were required to have the following characteristics: (1) maximum 1–2 nuclei per cell; (2) each subcellular compartment (nucleus, membrane and cytoplasm) area had to have >10 pixels; and (3) nuclei in each round of staining have to have perfect alignment to the first round of staining (automatic tissue quality index = 1 at each round, which is the correlation between DAPI in each image). After the quality control steps, exposure time was re-set at 200 ms for cy3, 500 ms for cy5, 50 ms for eGFP and 20 ms for DAPI across all slides and markers. Finally, biomarker intensities were log2 transformed. After all the preprocessing and quality control steps (and exclusion of the control slides), we have 379 408 epithelium cells of 13 cases for all the markers with available valid cell-level intensities.

For immune cell analysis (CD4, CD8 and CD20), support vector machine classification was used to classify immune cells. Model training was carried out by manual annotations using a tool developed in 'FIJI' program to annotate CD4, CD8 and CD20-positive cells for ground truth and model training. These features were used to assign each cell in the image, a probability of being a member of the CD4+, CD8+ or CD20+ classes. Ultimately, each cell in the image was assigned to one of the following classes: negative for all immune markers; cytotoxic T cells (CD8+); T-helper cells (CD4+) and B cells (CD20+).

Cellular Heterogeneity and Shannon Index Methods

Shannon diversity index was applied to the cell data to measure the diversity of cells within the data set using the proportion of each cell types. It is a quantitative index that was originally developed by Shannon³¹ to quantify uncertainty in strings of text. It has been applied to ecology applications to quantify to proportion of individuals belonging to a species in a given data set, and quantifies the uncertainty in predicting the species identity of an individual taken at random from the data set. It is calculated as the following:

$$H = - \sum_{i=1}^R p_i \ln p_i$$

where R is the total number of clusters and p_i is the proportion of the cluster i . The higher H is, the more heterogeneity is present in the data set. For example, in the case of this study, if all the cells from the same duct are assigned to one K-mean cluster, the H value approaches to zero. However, if all the cells come from multiple clusters evenly, the H value is high and approaches to $\ln R = \ln(8) = 2.08$ ($R = 8$ clusters in our dataset). A Shannon index was computed for each duct, then for a single patient, the average and s.d. of Shannon index from all their

imaged ducts was calculated to describe duct-to-duct diversity.

K-means clustering algorithm was applied to all cellular data (379 408 cells for all 13 patients) in *R* for $K=2,3, \dots, 15$ and repeat 10 runs with randomly selected initial K centroids for each K . To help determine the best K , we used Consensus Clustering (*R* library ConsensusClusterPlus) and based on the delta area metric. The best results were obtained with

a cluster solution using the following seven epithelial markers: EGFR, HER2, HER4, SLC7A5, pmTOR, CD44v6 and CD10. A heatmap was plotted to show cluster definition. Using cell IDs and spatial coordinates, the epithelial cell clusters were then overlaid with the tissue images using a color-coding schema.

Once the clusters were defined at cell level, the compositions of clusters for each patient and tumor field of views was calculated using level enrichment

a Region Selection

b Data Collection

c Single Cell Analysis

Figure 1 Data collection and analysis workflow. (a) Tissue sections are scanned for autofluorescence and DAPI, and an H&E-like image is generated allowing for specific regions to be analyzed by multiplex IF. (b) Enlarged region from H&E scan and a companion composite image of multiplex staining for epithelial/tumor cells (anti-pan cytokeratins AE1 and pck26), CD3 T cells, and CD20 B cells are shown. (c) Output images from single-cell analysis workflow showing (left) separation of tumor (red) and stroma (green), (middle) single-cell segmentation (blue—nuclei, red—epithelial membranes, green—epithelial cytoplasm), and (right) mapping of cell expression patterns based on K-means clustering back on original images.

statistics. Each patient was represented by a vector of eight elements with each element being the percentage of cells falling in that cluster and all the elements adding up to 100 percent. A pie chart was also generated to visually assess the cell type variation across all patients. The same statistic was also calculated for each field of view and the relationship between enrichment and the histological scores at patient level was assessed.

For the immune markers (CD4, CD8, CD20), we limited analysis to the stromal region (543 355 cells in total) as described above. The patient-level immune profiles were then summarized by the proportion of CD4 positive cells relative to stromal cells, the proportion of CD8-positive cells, the proportion of CD20-positive cells, the total T cells (CD4+CD8), the CD4 to CD8 ratios, and the T-cell to B-cell ratios. The relationship between immune profiles and traditional histological scores at patient level was analyzed. The correlation between epithelium cell clusters and stromal area immune positivity was also analyzed.

Results

Heterogeneity within Ductal Carcinoma *In Situ*—Single Duct Analyses within Same Patients

To evaluate expression heterogeneity of marker expression in ductal carcinoma *in situ*, we analyzed 416 individual ducts involving ductal carcinoma *in situ* on 13 patient tissue sections (average 29 ducts per patient; range: 20–40; see Figure 1 and Supplementary Table S2 for work-flow details) containing a total of 379 408 epithelial cells from 13 cases (average 29 185 cells per patient). Similarly, 543 355 stromal cells (average 41 796 cells per patient) were studied for expression of the three immune markers CD4, CD8 and CD20. Using this cellular data, cluster analysis for epithelial markers was performed within ducts, between ducts and between patients. Additional correlation analysis was performed to investigate relationships between these epithelial clusters and immune cell infiltration.

Analysis of the markers in the epithelial compartment showed heterogeneity of expression within single individual ducts from the same patient

Figure 2 Evaluation of heterogeneity in antibody staining for individual markers. (a) Representative composite images of staining from a typical ductal carcinoma *in situ* lesion (left), and a region of normal ducts from the same patient sample. (b) Boxplot of HER4, HER2 and pmTOR intensity for all the cells of one example patient. X axis is the ducts of the lesion. Y axis is the log₂(protein intensity). Within each duct, each point is a cell, the black dot is the median cell intensity and the boxplot around the black dot is the 25th and 75th quantiles. This figure shows the cell-to-cell and duct-to-duct variabilities for HER2, HER4 and pmTOR.

Figure 3 Determining complex patterns of protein expression in ductal carcinoma *in situ*. (a) K-means clustering was applied to single-cell data to determine patterns of expression for seven markers. Each cluster group represents a pattern of unique cellular expression. Scale set with blue representing low/negative expression and red is highest expression. (b) Cluster IDs mapped back onto images to confirm accuracy of segmentation demonstrates variable levels of heterogeneity. Shown are three representative cases with high, intermediate and low heterogeneity. Colors of cells correspond to color code given to cluster groups in Figure 3a. (c) Boxplot of cellular heterogeneity of patients. X axis is the eight cell-level clusters. Y axis is the percentage of cells within each cluster. Each point is a duct. Most of the ducts for case 5 has > 80% cells in cluster 4. Case 10 has majorly clusters 1 and 5. However, case 13 has three main clusters and the percentage of cells within each duct can vary from 0 to 100% for cluster 4, 0 to 90% for cluster 5 and 0 to 60% for cluster 6.

(Figures 2a and b; Supplementary Figures S1A–C). Expression pattern for markers such as HER4 and HER2 was relatively stable across multiple cells/ducts. On the other hand, pmTOR, CD44v6 and CD10 were infrequently expressed (5%, 2% and 2%, respectively) and varied significantly among cells from the same duct (Supplementary Figure S2). To better understand marker co-expression and quantify heterogeneity, a clustering algorithm (K-means) was applied to the epithelial cell biomarker data. Eight different clusters were identified based on expression of seven epithelial markers (Figure 3a); these cluster IDs were mapped back to the cells as illustrated as in Figure 3b (Supplementary Table S3). The proportion of each cluster was quantified and heterogeneity was assessed using Shannon diversity index. Briefly, clusters 2 and 8 both showed high expression of CD10 and EGFR, but cluster 8 showed low expression of other markers. HER2 expression was high in clusters 3 and 7; pmTOR expression was higher in clusters 1–3 and 5. There was moderate heterogeneity in distribution of eight clusters within each duct (average Shannon index 0.76; range 0–1.61).

Heterogeneity within Ductal Carcinoma *In Situ* – Single Patient Analyses

To further address the issue of the heterogeneity of ductal carcinoma *in situ* lesions within the same

patient (Figure 3c and Supplementary Figures S3–S10), the distribution of the clusters were analyzed in different ducts from the same patient. Four patients had a dominant cellular phenotype (>80%) represented by cluster groups 3 and 7 (high HER2), cluster 6 (high HER4 and SLC7A5 and low HER2 and pmTOR) or cluster 4 (no dominant markers define this group). In five patients, the pattern was more heterogeneous with 50–70% of the cells belonging to cluster 1 (moderate to high levels for all markers except EGFR and CD10). The remaining patients had a strong representation of clusters 4 and 5 (CD44v6 and pmTOR) cells. Within each patient, the average Shannon index across all ducts ranged from 0.33 to 1.02 (s.d. 0.09–0.38). These data document the presence of high degree of heterogeneity at the protein level in patients with ductal carcinoma *in situ*. Additional analyses were performed to estimate the minimal number of ducts required to represent all eight clusters. A cluster was deemed to be represented, if >5% of the duct cells belong to the particular cluster. On average, the number of clusters represented within a single duct was 2.5 (range 1–6), which corresponds to Shannon index 0.76. As the distribution of clusters within ducts is uneven, the analysis of eight ducts would be sufficient to represent all eight clusters, that is, within- and between-duct heterogeneity.

Figure 4 Immune quantification in ductal carcinoma *in situ* samples. (a) Representative images from three cases of ductal carcinoma *in situ* with differing levels of local immune response. Tumor cells (turquoise), CD20 (red), CD4 (green) and CD8 (blue). (b) Boxplot of immune marker enrichment of patients. X axis is the four main immune types: CD4, CD8, CD20 and total T cells (CD4+CD8). Y axis is the proportion of positive immune cells. Each point is a duct. Most of the ducts for case 11 have very few immune cells. Case 6 has moderate level of immune cells. However, case 1 has relatively higher immune proportions for both T cells and B cells.

Correlation of Stromal/Immune Analysis with Epithelial Expression Pattern

To analyze the impact of heterogeneity on immune infiltration, immune cells were quantified in both the stromal and epithelial compartments (representative data in Figures 4a and b). The stromal tumor-infiltrating lymphocytes were classified as (CD20+) B cells or (CD4+ or CD8+) T cells. The distribution of stromal tumor-infiltrating lymphocytes in ductal

carcinoma *in situ* cases (Supplementary Table S4) was either a mixed B-cell (CD20+) or T-cell response ($n = 10$), or one dominated by CD4 T cells ($n = 1$). The relationship between histology and immune markers was assessed in each field of view for all patients (Figures 5a-c). Although the sample size is small, larger lesion size was positively correlated with proportion of cluster 7 (high HER2) cells and CD20 infiltration ($P = 0.002$ and 0.0027 from *t*-test,

Figure 5 Mapping of epithelial and immune components in ductal carcinoma *in situ*. (a) Representative pseudo-H&E images showing four cases of ductal carcinoma *in situ* with local immune responses. (b) Composite immunofluorescence images for fields shown in a. Markers include HER4, SLC7A5, HER2, pmTOR, CD44v6, CD10 and EGFR. (c) Cluster IDs mapped back to images allow clear visualization of cells with different patterns of expression. Colors correspond to cluster IDs in Figure 3, and pie charts shown in inset are for the specific field of view—the inner pie shows relative abundance of the eight epithelial groups, and the outer ring show proportions of different immune populations. (d) Verification of immune patterns after semi-automated quantification. CD20-positive cells in red, CD4 in green, and CD8 in blue.

respectively). Higher nuclear grade was associated with proportion of epithelial cluster 5 (high CD44v6 and pmTOR) cells (P -value 0.008) and CD20 positivity ($P=0.019$; F-test). The cases were further analyzed as per four subtypes defined by expression of ER, PR and HER2 (ER+/PR-/HER2-, ER+/PR+/HER2-, ER+ HER2+, and HER2+) accounting for 3, 7, 1 and 2 cases, respectively. Proportion of cluster 3 cells was lower in all ER+PR+ cases ($P=0.05$ from F-test). The ER+/HER2- cases had lower representation of cluster 7 cells ($P=0.08$ from F-test). The HER2+ patients had higher B cells irrespective of ER status (F-test; $P=0.003$).

A composite image representing the relationship of the epithelial clusters with immune markers is illustrated in Figure 6. This clearly highlights the degree of heterogeneity in protein expression in ductal carcinoma *in situ*. Cluster 2 (high EGFR and CD10 expression) was associated with a high T-cell to B-cell ratio (Spearman correlation coefficient=0.83, $P=0.0004$), whereas cluster 7 (high HER2 expression) was associated with a B-cell response (Spearman correlation coefficient=0.68, $P=0.009$). The relative scarcity of intra-epithelial tumor-infiltrating lymphocytes made correlations difficult.

Figure 6 Summary of epithelial and associated immune patterns in ductal carcinoma *in situ*. Pie charts for cluster distribution and immune profile for each patient. The inner pie is composed of the eight epithelial cell clusters. The outer ring is composed of the immune cell types as a proportion of the stromal space.

Discussion

Tumors being monoclonal in origin were believed to lack significant cellular diversity. This was particularly true for ductal carcinoma *in situ*, which is thought to arise in/from terminal duct units and then spread outward to involve the ductal system in a segmental manner. The currently used CAP classification system is based on the finding of uniformity of nuclear (grade) morphology in ductal carcinoma *in situ* lesions.³² These factors feed in to the notion that ductal carcinoma *in situ* is a homogeneous lesion. Recent studies have documented intra-tumor heterogeneity in several cancers as an important determinant for tumor recurrence and resistance to therapy.³³ The extent of intra- and inter-ductal heterogeneity in ductal carcinoma *in situ* has not been yet well documented.

The traditional methods of assessing heterogeneity involve studying different regions from the same lesion, extracting DNA and studying the mutational profile. This approach is excellent for larger lesions and fresh tumors but has limited utility in ductal carcinoma *in situ*, where the lesions are small, and admixed with normal tissues. Furthermore, these methods only permit regional assessment of heterogeneity and do not permit single-cell analysis of lesion diagnosed only after histological examination. Multiplexed immunofluorescent staining and single-cell analysis of cancer tissue provides a potentially powerful approach for measuring tumor cell expression and heterogeneity in intact formalin-fixed paraffin-embedded tissue, avoiding the 'grind and find' approach required for genomic analyses and potential loss of spatial information.¹⁰ Previous work using this multiplexed immunofluorescence measured 27 proteins at single-cell level in 26 patients with untreated stage II/III breast cancer and found eight clusters of proteins that were co-expressed within the tumor region. The ER-negative cluster phenotype was significantly correlated with uptake of radiotracer 18F-fluorodeoxyglucose.³⁴ Quantitative intra-tumor spatial heterogeneity algorithms have been employed in multiplexed invasive breast cancer.³⁵ Similarly, a nine multi-marker analysis of Reed Sternberg cells has been used for diagnosis of Hodgkin lymphoma.³⁶ Other reports have showed how multiplexing could decipher the complexity of the tumor-immune relationships and patient outcomes. Multispectral imaging of CD3, CD8, FoxP3, CD163, PD-L1 was used to analyze the tumor microenvironment in melanoma patients and predicted ability to generate tumor-infiltrating lymphocytes for T-cell therapy.³⁷

We have therefore applied a multiplex immunofluorescence based approach to study the heterogeneity of ductal carcinoma *in situ* of protein expression pattern in formalin-fixed paraffin-embedded tissues. This analysis is performed on a single 5 μ m section and represents a paradigm shift in that it permits multiplex *in situ* protein analysis at

the single-cell level. Furthermore, co-expression analysis, application of clustering algorithms and annotation of the expression data to single cells are possible, whereas retaining the tissue structural context.

Co-expression and correlation analysis of seven different proteins of epithelial and three immune origin showed marked differences in expression of proteins in the cells. A single duct was found to contain cells located adjacent to one another expressing different proteins. The application of K-means clustering identified distinct co-expression profiles within individual cells of the ductal carcinoma *in situ* lesions. The use of these algorithms, although typical for gene expression studies, has not been commonly applied for tissue based protein analysis. This analysis identified moderate to marked degree of heterogeneity of protein expression by cells within a single duct (average Shannon index 0.76; range 0–1.61). This highlights the degree of heterogeneity that is not appreciated by mutational analyses. More importantly, it raises the possibility of co-operative action, as a colony, in which single cells could have different roles in propagating the growth and progression of the disease.

Further analysis of multiple ducts from the same patient showed moderate degree of heterogeneity with individual ducts being composed of one or 2 (co-) dominant clusters/clones or showing a scattered mix of phenotypes. Similar moderate degree of heterogeneity was seen in ducts from different patients. Higher expression of cluster 5 cells was associated in nuclear grade 2 ($P=0.008$; F-test), particularly in ER+ patients. There was also a trend for lower cluster 3 and 7 expression in ER+ patients. Grade and ER/PR expression have prognostic value in determining the likelihood of recurrence and/or progression to invasive cancer. One could postulate that the observed tumor heterogeneity could have a bearing on the propensity to recur or progress.

In the current study, specific protein expression profiles were associated with TIL responses. In invasive breast cancer, immune cells have been documented to contribute to prognosis even in cases treated with chemotherapy.^{27,28,38–40} The amount of tumor-infiltrating lymphocytes has been correlated with mutational burden in gastrointestinal cancers.⁴¹ In breast cancer, specific mutational signatures have been associated with tumor-infiltrating lymphocytes.⁴² Tantalizingly, an association of epithelial clustering pattern with the presence and type of immune infiltrates was observed. This analysis provides the first evidence of that simultaneous analysis of both the epithelial and immune/stromal components might be necessary to understand the complex milieu in ductal carcinoma *in situ* lesions.

The current study is based on the analysis of a larger number of cells and ducts albeit from a limited number of patients with no outcome data (work is

ongoing to expand sample size with outcomes and will include spatial cell analysis for deeper analysis of cell-to-cell interactions and relationships). In spite of these limitations, the study raises several issues that are fundamental to understanding of the biology of cancer, particularly breast cancer. The study, for the first time, documents the significant heterogeneity in protein expression in pre-invasive breast cancer. The documentation of moderate tumor heterogeneity within cells of a single duct has profound implications. It seems to suggest that even at the pre-invasive stage, cancer is a heterogeneous disease, and perhaps different clones fighting for dominance. This is consistent with the stochastic model of cancer evolution.^{43,44} On the other hand, the finding of single dominant clone in some cases would support the 'Big Bang' model.⁴⁵ Does this indicate one should not focus on analyzing mass lesions for mutational and protein expression and focus on single cells? Is this the future of single-cell sequencing? If so, how many cells need to be sequenced is an important question. From the patient perspective, 'Is heterogeneity or lack of it good?' Homogeneous lesions could indicate the establishment of clonal dominance as a sign of aggressiveness. On the other hand, as cases with marked heterogeneity would have different clones, one or more of them could survive/give rise to invasive disease. Understanding the biological impact is complicated by the fact that it is rare to see direct progression of ductal carcinoma *in situ* to invasive carcinoma and the low recurrence rates in treated patients. The only way forward is to analyze a large series of well-annotated cases with established follow-up or a case-control design study. The current study provides the scientific basis for the design and performance of analysis of well-annotated ductal carcinoma *in situ* patient cohorts using multiplexed immunofluorescence technique and clustering algorithm described herein. These will lead to identification of novel clustering patterns that are related with outcomes.

Acknowledgments

This study is supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA194600 to SS Badve and MJ Gerdes and internal funds from General Electric to F Ginty and MJ Gerdes. In addition, SS Badve is supported by 100 Voices of Hope and Joshua Edwards Endowed Chair. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

- Gerlinger M, Rowan AJ, Horswell S, *et al*. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 2012;366:883–892.
- Govindan R. Cancer. Attack of the clones. *Science* 2014;346:169–170.
- de Bruin EC, McGranahan N, Mitter R, *et al*. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* 2014;346:251–256.
- Zhang J, Fujimoto J, Zhang J, *et al*. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256–259.
- Navin NE. Cancer genomics: one cell at a time. *Genome Biol* 2014;15:452.
- Smith BD. When is good enough really good enough? Defining the role of radiation in low-risk ductal carcinoma *in situ*. *J Clin Oncol* 2015;33:686–691.
- Solin LJ, Gray R, Hughes LL, *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* 2015;33:3938–3944.
- Solin LJ, Gray R, Baehner FL, *et al*. A multigene expression assay to predict local recurrence risk for ductal carcinoma *in situ* of the breast. *J Natl Cancer Inst* 2013;105:701–710.
- Badve S, A'Hern RP, Ward AM, *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* 1998;29:915–923.
- Gerdes MJ, Sevinsky CJ, Sood A, *et al*. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc Natl Acad Sci USA* 2013;110:11982–11987.
- Zhou W, Johansson C, Jirstrom K, *et al*. A comparison of tumor biology in primary ductal carcinoma *in situ* recurring as invasive carcinoma versus a new *in situ*. *Int J Breast Cancer* 2013;2013:582134.
- Pape-Zambito D, Jiang Z, Wu H, *et al*. Identifying a highly-aggressive DCIS subgroup by studying intra-individual DCIS heterogeneity among invasive breast cancer patients. *PLoS ONE* 2014;9:e100488.
- Lebeau A, Unholzer A, Amann G, *et al*. EGFR, HER-2/neu, cyclin D1, p21 and p53 in correlation to cell proliferation and steroid hormone receptor status in ductal carcinoma *in situ* of the breast. *Breast Cancer Res Treat* 2003;79:187–198.
- Lari SA, Kuerer HM. Biological markers in DCIS and risk of breast recurrence: a systematic review. *J Cancer* 2011;2:232–261.
- Collins LC, Schnitt SJ. HER2 protein overexpression in estrogen receptor-positive ductal carcinoma *in situ* of the breast: frequency and implications for tamoxifen therapy. *Mod Pathol* 2005;18:615–620.
- Barnes NL, Khavari S, Boland GP, *et al*. Absence of HER4 expression predicts recurrence of ductal carcinoma *in situ* of the breast. *Clin Cancer Res* 2005;11:2163–2168.
- Altintas S, Lambein K, Huizing MT, *et al*. Prognostic significance of oncogenic markers in ductal carcinoma *in situ* of the breast: a clinicopathologic study. *Breast J* 2009;15:120–132.
- Buckley N, Boyle D, McArt D, *et al*. Molecular classification of non-invasive breast lesions for

- personalised therapy and chemoprevention. *Oncotarget* 2015;6:43244–43254.
- 19 Afify A, McNiel MA, Braggins J, *et al*. Expression of CD44s, CD44v6, and hyaluronan across the spectrum of normal-hyperplasia-carcinoma in breast. *Appl Immunohistochem Mol Morphol* 2008;16:121–127.
 - 20 Miller TJ, McCoy MJ, Hemmings C, *et al*. Objective analysis of cancer stem cell marker expression using immunohistochemistry. *Pathology* 2017;49:24–29.
 - 21 Vermeulen JF, van Brussel AS, Adams A, *et al*. Near-infrared fluorescence molecular imaging of ductal carcinoma *in situ* with CD44v6-specific antibodies in mice: a preclinical study. *Mol Imaging Biol* 2013;15:290–298.
 - 22 Juhasz C, Nahleh Z, Zitron I, *et al*. Tryptophan metabolism in breast cancers: molecular imaging and immunohistochemistry studies. *Nucl Med Biol* 2012;39:926–932.
 - 23 Hayashi K, Anzai N. Novel therapeutic approaches targeting L-type amino acid transporters for cancer treatment. *World J Gastrointest Oncol* 2017;9:21–29.
 - 24 Moritani S, Kushima R, Sugihara H, *et al*. Availability of CD10 immunohistochemistry as a marker of breast myoepithelial cells on paraffin sections. *Mod Pathol* 2002;15:397–405.
 - 25 Stanton SE, Disis ML. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *J Immunother Cancer* 2016;4:59.
 - 26 Denkert C, Loibl S, Noske A, *et al*. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 2010;28:105–113.
 - 27 Perez EA, Ballman KV, Tenner KS, *et al*. Association of stromal tumor-infiltrating lymphocytes with recurrence-free survival in the N9831 adjuvant trial in patients with early-stage HER2-positive breast cancer. *JAMA Oncol* 2016;2:56–64.
 - 28 Adams S, Gray RJ, Demaria S, *et al*. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol* 2014;32:2959–2966.
 - 29 Hammond ME, Hayes DF, Dowsett M, *et al*. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Arch Pathol Lab Med* 2010;134:907–922.
 - 30 Wolff AC, Hammond ME, Hicks DG, *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013;31:3997–4013.
 - 31 Shannon CE. A mathematical theory of communication. *Bell Syst Tech J* 1948;379–423.
 - 32 Lester SC, Connolly JL, Amin MB. College of American Pathologists protocol for the reporting of ductal carcinoma *in situ*. *Arch Pathol Lab Med* 2009;133:13–14.
 - 33 Badve S, Gokmen-Polar Y. Tumor heterogeneity in breast cancer. *Adv Anat Pathol* 2015;22:294–302.
 - 34 Sood A, Miller AM, Brogi E, *et al*. Multiplexed immunofluorescence delineates proteomic cancer cell states associated with metabolism. *JCI Insight* 2016;1:e87030.
 - 35 Spagnolo DM, Gyanchandani R, Al-Kofahi Y, *et al*. Pointwise mutual information quantifies intratumor heterogeneity in tissue sections labeled with multiple fluorescent biomarkers. *J Pathol Inform* 2016;7:47.
 - 36 Hollman-Hewgley D, Lazare M, Bordwell A, *et al*. A single slide multiplex assay for the evaluation of classical Hodgkin lymphoma. *Am J Surg Pathol* 2014;38:1193–1202.
 - 37 Tsujikawa T, Kumar S, Borkar RN, *et al*. Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis. *Cell Rep* 2017;19:203–217.
 - 38 Adams S, Goldstein LJ, Sparano JA, *et al*. Tumor infiltrating lymphocytes (TILs) improve prognosis in patients with triple negative breast cancer (TNBC). *Oncoimmunology* 2015;4:e985930.
 - 39 Salgado R, Denkert C, Demaria S, *et al*. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol* 2015;26:259–271.
 - 40 Loi S, Michiels S, Salgado R, *et al*. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol* 2014;25:1544–1550.
 - 41 Giannakis M, Mu XJ, Shukla SA, *et al*. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. *Cell Rep* 2016;17:1206.
 - 42 Smid M, Rodriguez-Gonzalez FG, Sieuwerts AM, *et al*. Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. *Nat Commun* 2016;7:12910.
 - 43 Yates LR, Campbell PJ. Evolution of the cancer genome. *Nat Rev Genet* 2012;13:795–806.
 - 44 Sidow A, Spies N. Concepts in solid tumor evolution. *Trends Genet* 2015;31:208–214.
 - 45 Sottoriva A, Kang H, Ma Z, *et al*. A Big Bang model of human colorectal tumor growth. *Nat Genet* 2015;47:209–216.

Supplementary Information accompanies the paper on Modern Pathology website (<http://www.nature.com/modpathol>)