Genomic Alterations during the In Situ to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System

Anne Trinh1,2,3, Carlos R. Gil Del Alcazar1,2,3, Sachet A. Shukla1,2,3,4, Koei Chin5,6, Young Hwan Chang5,6, Guillaume Thibault5, Jennifer Eng5, Bojana Jovanović1,2,3,4, C. Marcelo Aldaz7, So Yeon Park8, Joon Jeong9, Catherine Wu1,2,3,4, Joe Gray5,6, and Kornelia Polyak1,2,3,4

ABSTRACT

The drivers of ductal carcinoma in situ (DCIS) to invasive ductal carcinoma (IDC) transition are poorly understood. Here, we conducted an integrated genomic, transcriptomic, and whole-slide image analysis to evaluate changes in copy-number profiles, mutational profiles, expression, neoantigen load, and topology in 6 cases of matched pure DCIS and recurrent IDC. We demonstrate through combined copy-number and mutational analysis that recurrent IDC can be genetically related to its pure DCIS despite long latency periods and therapeutic interventions. Immune “hot” and “cold” tumors can arise as early as DCIS and are subtype-specific. Topologic analysis showed a similar degree of pan-leukocyte-tumor mixing in both DCIS and IDC but differ when assessing specific immune subpopulations such as CD4 T cells and CD68 macrophages. Tumor-specific copy-number aberrations in MHC-I presentation machinery and losses in 3p, 4q, and 5p are associated with differences in immune signaling in estrogen receptor (ER)-negative IDC. Common oncogenic hotspot mutations in genes including TP53 and PIK3CA are predicted to be neoantigens yet are paradoxically conserved during the DCIS-to-IDC transition, and are associated with differences in immune signaling. We highlight both tumor and immune-specific changes in the transition of pure DCIS to IDC, including genetic changes in tumor cells that may have a role in modulating immune function and assist in immune escape, driving the transition to IDC.

Implications: We demonstrate that the in situ to IDC evolutionary bottleneck is shaped by both tumor and immune cells.

Introduction

Ductal carcinoma in situ (DCIS) of the breast is a preinvasive cancer characterized by abnormal proliferation of neoplastic epithelial cells confined within the mammary ductal-lobular unit and separated from the stroma by an intact myoepithelial cell layer and basement membrane (1). While over 50,000 cases of DCIS are diagnosed each year in the United States, only one third of these will progress to invasive ductal carcinoma (IDC) within 30 years, presenting a natural evolutionary bottleneck (1). Accurate prediction of the likelihood of progression would alleviate over-treatment of the disease. However, achieving this is hampered by our relatively limited knowledge of the molecular processes underlying the DCIS-to-IDC transition.

Characterizing determinants of progression remains challenging due to difficulties with acquiring matched DCIS-IDC samples from treatment-naïve patients and long latency periods after the initial diagnosis of pure DCIS. It is estimated that 10% of patients will relapse with invasive disease after 10 years, with ipsilateral recurrences at least twice as common as contralateral recurrence (2). This trend of local recurrence suggests that the new lesion is genetically related to the prior lesion, and studies of matched pure DCIS and IDC may be informative in uncovering driver events.

Genetic studies of DCIS and IDC cases have shown limited stage-specific differences that could predict progression. Whole-exome sequencing (WES) of DCIS and IDC has shown high similarity in copy-number profiles between the two histologic entities, with notable copy-number aberrations (CNA) including 1q, 8q, 17q, 20q amplification and 8p, 11q, 16q, 17p loss (3–6). Activation of oncogenes including PIK3CA, GATA3, ERBB2, and loss of tumor suppressors TP53 and CDKN2A, are thought to be early breast cancer driver events due to their similar frequencies in DCIS and IDC (5, 7). Coordinate selection of multiple genes in chromosome scale aberrations that provide low-level fitness advantages has also been implicated as a survival mechanism (8). High genetic similarity has been noted in studies of matched synchronous DCIS and IDC (4, 9–12); however, these are reflective of a timepoint beyond the evolutionary bottleneck and the full repertoire of mediators of the successful transition from DCIS to IDC cannot be established from these studies.

In contrast to tumor epithelial cells and genetic changes, multiple cell types composing the tumor microenvironment show tumor progression stage-specific gene expression and epigenetic differences (13, 14). Genes involved in extracellular matrix (ECM)
organization including collagens, matrix metalloproteinases (MMP), and cell adhesion molecules are upregulated in IDC compared with DCIS (15, 16). Differences in immune cells including T, B, and natural killer (NK) cells have also been noted between DCIS and IDC (16, 17), and interleukin (IL4, IL12, IL23) signaling is differentially expressed in DCIS compared with normal breast tissue with changes maintained in IDC (17). These studies suggest that the tumor microenvironment and immune system play major roles in the DCIS to IDC transition.

Immuno-surveillance exerts constant selective pressure against tumor cells, which may lead to intratumor heterogeneity observed in DCIS (18). We have previously demonstrated that in HER2+ and triple-negative (TN) DCIS there is an activated immune environment characterized by higher frequency of cytotoxic CD8+ T cells compared with IDC (19). In contrast, the microenvironment of TN IDC is often immunosuppressive with higher PD-L1 expression in tumor epithelial cells and higher expression of regulatory T cell (Treg) markers (19). Although PD-L1 expression by DCIS tumor epithelial cells is rare (19, 20), high proportions of PD-L1+ tumor-infiltrating lymphocytes (TIL), Tregs, and CD68+ macrophages have been observed in high-grade DCIS (20, 21), implying immune-editing (22).

We hypothesize that the DCIS to IDC transition is mediated by concurrent tumor-specific and immune microenvironmental changes. To explore this, here we describe, for the first time, integrated genetic, gene expression and topologic analysis of a small cohort of matched pure DCIS and recurrent IDC, focusing on immune-related alterations including changes in predicted neoantigens.

Materials and Methods

Patient cohort
Patient material was collected from Yonsei University and Seoul National University, Korea between 2000 and 2014. All cases were reviewed by expert pathologists (S.Y. Park and J. Jeong) and scored for estrogen receptor (ER), progesterone receptor (PR), and HER2 status. Written informed consent was obtained from all patients, and all protocols used were approved by the ethical and institutional review boards, in concordance with the declaration of Helsinki. The patient protocols used were approved by the ethical and institutional review boards.

Exome and RNA sequencing
Patient samples were macrodissected into stromal and epithelial fractions from FFPE (formalin fixed paraffin embedded) tissue slides, with examples of segmented regions highlighted in Supplementary Fig. S1B. DNA and RNA were extracted from epithelial fractions from matched DCIS and IDC cases and corresponding normal breast using AllPrep DNA/RNA FFPE Kit (Qiagen, catalog no. 80234). Exome and RNA libraries were prepared by the Genomic Platform at the Broad Institute using Nextera Rapid Capture Exome Kit (Illumina, catalog no. FC-140-1000) and TruSeq RNA Access Library Prep kit (Illumina, catalog no. 2002189) and sequenced on HiSeq4000 and Hiseq2000 (SBS flow cells, respectively. Reads were aligned to the hg37 genome using bwa (RRID:SCR_010910) for exome-data (23) and STAR (RRID:SCR_015899; ref. 24) for RNA-sequencing data. Further details are described in Supplementary Methods and exome coverage is summarized in Supplementary Fig. S2.

Histology and cyclic immunofluorescence
Whole-tissue mounts were sectioned to a thickness of 4 μm and stained with hematoxylin and eosin. The cyclic immunofluorescence (cycIF) method was applied to cases 8 and 9 to create highly multiplexed images as described previously (25), and further details including antibodies are described in Supplementary Methods and Supplementary Table S2. Ten cycles of immunofluorescence using 4 antibodies of interest and DAPI, imaging and fluorophore bleaching was performed. Collected images are then registered based on DAPI staining, and image segmentation and feature extraction were performed. Intensities within a tissue were scaled to a [0, 1] distribution and phonograph (RRID:SCR_016919; ref. 26) was used to classify cells into clusters.

Statistical analysis
Code to reproduce figures and statistical calculations in this study is available at https://github.com/polynk-lab/matchedDCISIDC and is described in Supplementary Methods.

WES
The Getz Lab CGA WES Characterization pipeline at the Broad Institute (https://docs.google.com/document/d/1VO2K_fhU0x 3mB9NjLjWZGSw94WbTepB3cB68/edit#heading=h-5yby87z2bcs) has been optimized for analysis of both frozen tissue and archival FFPE. Segmented contiguous chromosomal regions were called using GATK (RRID:SCR_001876) CNV (27). A log fold change of ±0.3 was used to define gains and losses from array CGH data from Lesurf and colleagues (16). Variant discovery was performed using a combination of MuTect1 (27), MuTect2 (27) and Strelka (RRID:SCR_005109; ref. 28) and was annotated using Oncotator (RRID:SCR_005183; ref. 29). Ploidy and variants in normal exome samples were called using GATK GermlineCNVCaller and HaploTypeCaller, respectively (27) and annotated for pathogenicity using ClinVar (RRID:SCR_006169; ref. 30).

Topology
Hematoxylin and eosin (H&E) images were digitally segmented using Qupath (RRID:SCR_018257; ref. 31) and cells were classified into epithelial, stromal, and immune cells. Spatial-pattern analysis was performed using k-nearest neighbor distances and Morisita-Horn index for spatial overlap (32).

Gene expression
RNA sequencing of cases 1, 5, 2DCIS, 3DCIS were excluded from analysis due to poor quality. Differential gene expression analysis was performed using DESeq2 (RRID:SCR_000154; ref. 33), and reported fold changes used in gene-set enrichment analysis was performed using HTSAnalyzer (34) using the c2 compendium (35, 36). Immune signature analysis was performed using single-sample GSEA (ssGSEA; RRID:SCR_003199) using signatures manually curated from the literature (19, 37). Proportions of different immune cell populations was inferred using a combination of TIMER (RRID:SCR_018737; ref. 38), CIBERSORT (RRID:SCR_016955; ref. 39), xCELL (40), and EPIC (41). For inferred fractions, statistical differences were assessed using a beta regression model, whereas enrichment scores were assessed using Wilcoxon-rank sum test.
Neoantigen prediction

All nonsynonymous mutations and frame shift mutations were in silico translated using SIFT (RRID:SCR_012813; ref 42) into peptides of length 8–11 amino acids. HLA-type was inferred from exome data using Polysolver (43). Neoantigens were predicted using NetMHC-Pan4 (RRID:SCR_018182; ref 44), and estimated binding affinity of <50 nM (strong) and <200 nM (weak) were used to predict potential binders. Neoantigens were validated as having at least 2 RNA-sequencing reads supporting the mutation.

BCR/TCR repertoire

CDR3 recombined regions in B cells and T cells were inferred from RNA using MixCR (RRID:SCR_018725) (45). CDR3 reads were filtered to those that have at least two supporting reads and have an amino acid sequence length of at least 6. These were searched against the IEDB database (46) of known CDR3 sequences.

Accession codes

RNA-Sequencing datasets have been deposited to GEO with accession number GSE144020.

Results

Copy-number aberrations are early events in the DCIS-to-IDC transition

We performed WES to identify differences in CNAs and mutational profiles between matched pure DCIS and recurrent IDC (Supplemental Figs. S1, S2A and S2B; Supplementary Table S1). Among all DCIS and IDC cases within our recurrence cohort, frequently observed changes included 1q, 8q, 16p, 17q amplification and 11q and 16q loss, consistent with previous observations in other DCIS datasets including the Abba and Lesurf cohorts (refs. 5, 16; Fig. 1A). For most DCIS-IDC pairs, copy-number profiles of the IDC closely resembled that of the DCIS, where on average 84% of bases had similar calls and Case 8 showed high correlation in copy-number ratios (Supplementary Fig. S2C).

In contrast to the other cases, 2 IDC exhibited aberrations in large contiguous chromosomal regions.

Focusing specifically on breast cancer–related oncopgenes, the most commonly observed gain across all cohorts was at the ERBB2 locus (Fig. 1B and C). 17q gain in IDC may drive weak HER2+ expression in the IDC. In contrast, 5 IDC retains this amplicon despite being characterized as ER−HER2+, and 3 IDC did not show 17q21 amplification despite being characterized as HER2+ by IHC and showed very weak RNA expression. ERBB2 amplifications in PAM50 HER2− cases were observed in other cohorts – the overall proportion of cases with ERBB2 amplification was 48% (14/29) and 40% (17/42) in the Abba and Lesurf cohorts. Other common gains included MYC (8q24) and FGFR1 (8p11). Loss of heterozygosity was observed at 16q22 (CDH1, CTCF) in all cohorts; however, gains were also noted in the Abba cohort. Losses were also noted in genes associated with DNA repair (BRCA2, BRCA1, TP53) and cell-cycle regulation (RB1).

Overall, the high genetic similarity between DCIS and IDC shown in our data combined with consistency in copy-number profiles with other DCIS cohorts (3, 4, 16) supports the early acquisition of copy-number changes (47).

Genetic overlap between pure DCIS and recurrent IDC

We assessed mutational burden in coding regions and found very little overlap both across patients and even within the same patient (Supplementary Table S3; Supplementary Fig. S3A and S3B). The total and coding mutational burden was similar in both DCIS and IDC, although a decrease was noted in cases 8 and 3 IDC and an increase in 5 IDC (Supplementary Fig. S3C). However, variant allele frequencies less than 0.08 cannot be reliably detected given the sample coverage (Supplementary Fig. S3D), and sampling bias, clonal expansion of a dominant clone, neutral drift, or bottleneck selection could explain the differences in these profiles.

Mutations in breast cancer–related genes shared between DCIS and IDC occurred in driver events including TP53 (cases 5 and 8), PIK3CA (case 9, E545K activation), and CDH1 (case 8) and RAD50 (case 8; Fig. 1C). The Abba cohort also showed frequent mutations in the known driver genes TP53 (17%), PIK3CA (20%), and GATA3 (14%; Fig. 1E). Other conserved coding mutations lie in genes associated with ubiquitin (DCAF16, UCHL5, USP11) and metabolism and mitochondrial activity (GGT7, GUSB, HADHB, SLC25A37; Supplementary Fig. S3B).

In cases, recurrent mutations have previously been identified in normal human tissues (48). Thus, we sought to determine whether such events occur in our cohort by profiling adjacent matched mammary normal tissue (Fig. 1D). Most samples harbored benign rsSNP variants for common oncopgenes including BRCA1/2 and KMT2C. However, we also found some variants that have been proposed to be risk factors or have uncertain clinical effect in the literature. A possible example is PIK3CA E545K in the known normal tissue and maintained in the DCIS (Supplementary Fig. S3E), indicating the possibility of a very early driver event.

Similarity in CNAs and the presence of conserved mutations points to a genetic relationship between pure DCIS and recurrent IDC despite long latency periods in cases 8 and 9. Case 5 had many shared mutations, suggesting contralateral seeding at the time of the DCIS. Cases 1 and 3 do not display mutational overlap and have quiet copy-number profiles. It is unclear whether they are directly genetically related. However, an ipsilateral recurrence only after 3 years in case 3 suggests the recurrence is likely to have stemmed from residual primary disease.

Immune-related changes between DCIS and IDC: differences in cellular composition and topology

Pathway enrichment analysis of genetic alterations appearing in all cohorts included those associated with cell cycle, DNA damage, apoptosis, and cancer-related pathways including RTK, ESR1, NOTCH, WNT signaling, which was also supported by RNA-sequencing data comparing normal breast tissue with DCIS in the Abba cohort (Fig. 1F; Supplementary Table S4). Moreover, also implicated were pathways pertaining to the extracellular matrix (ECM) and the immune system with cytokine signaling (the IL13/IL4 pathway supporting Th2 differentiation), innate immunity, and antigen processing and presentation. Both genetic and RNA profiling implied that the immune system has a major impact on the DCIS-to-IDC transition. Thus, we investigated immune-related changes in further detail.

We quantified tumor infiltrating lymphocytes (TILs) in whole-slide H&E images (Fig. 2A; Supplementary Fig. S1C) and assessed the spatial colocalization of tumor and immune cells using three metrics: (i) The interacting fraction, that is, the proportion of TILs within 10 μm of a tumor epithelial cell and capable of direct interaction (Fig. 2B); (ii) the average distance of the 3 closest neighbors of cell type x to each TIL (KNN, k = 3) and assessed the proportion within 50 μm (Fig. 2C); and (iii) the Morisita–Horn (MH) index of spatial correlation of two cells types (32), which provides a global metric for mixing independent of composition. To take into account differences in tumor
Figure 1.
Copy-number alterations and somatic mutations in the DCIS-to-IDC transition. **A**, Genome-wide summary of proportion of patients with observed CNAs in the Recurrence (our data), Abba (5), and Lesurf (16) cohorts. A z-score of 2 in GATK CNV was set to call gains and losses in the recurrence and Abba cohorts (exome-seq), and a threshold of \( \leq 0.3 \) in the Lesurf set (aCGH). Both DCIS and IDC samples are shown in the recurrence cohort. **B**, Summary of CNAs in breast cancer–associated oncogenes and tumor suppressors in the recurrence cohort. **C**, Summary of cancer-related CNAs and coding mutations in the recurrence cohort, and corresponding gene expression profiles. **D**, Variants found in adjacent normal mammary tissue. **E**, Summary of cancer-related CNAs and mutations in the Abba and Lesurf cohorts grouped by PAM50 subtype. **F**, Pathways implicated by genetic changes in all three cohorts and by RNA expression in the Abba cohort comparing DCIS to normal. Circle size reflective of the average enrichment score and line width reflective of the number of common genes in two pathways.
architecture between DCIS and IDC, we omitted regions containing only tumor cells when computing MH indices (Fig. 2D; Supplementary Methods).

IDC had either lower or the same fraction of immune cells compared with DCIS with the exception of case 1, which was a lymph node recurrence (paired one-tailed t test, \( P = 0.045 \)). No difference was observed for stromal or tumor cells (Supplementary Fig. S4A). Using the interacting fraction and kNN analysis, we did not detect a difference between DCIS and IDC (Fig. 2B and C), suggesting that there is a similar capacity for immune cells to interact with tumor cells in both DCIS and IDC. However, our kNN analysis highlighted that immune cells tended to cooccur with both immune cells and stromal cells in DCIS compared with IDC (Fig. 2C), which is evident in tertiary lymphoid structures found in cases 2 and 8 (Supplementary Fig. S1B).

This difference in spatial patterns could be attributed to lower TIL frequencies in IDC, which can be taken into account using the MH index. The MH-index for tumor-lymphocyte mixing was lower in all neoplastic lesions compared with adjacent normal breast tissue, indicating that some degree of compartmentalization and reduced immune-epithelial interaction occurs during tumorigenesis (Fig. 2D). Furthermore, once normalized for TIL content, there was less immune-stromal mixing in DCIS compared with normal tissue and IDC.

The analyses show that in matched pure DCIS and IDC the immune fraction in DCIS is comparable or higher. Furthermore, while compartmentalization occurs and there is less interaction between the three different cell types in neoplastic lesions, TIL-tumor spatial distances are comparable in DCIS and IDC.

**Immune hot and cold tumors arise in DCIS**

To understand differences in DCIS compared to IDC, we performed GSEA across all DCIS and IDC samples and conducted paired analysis for cases 8 and 9 (Fig. 3A; Supplementary Table S5). Common pathways enriched in DCIS include E-cadherin stabilization and ATM pathways, whereas WNT signaling was enriched in IDC, consistent with our genomic analysis. Sample 8 was enriched for cell-cycle markers in IDC, indicating higher proliferation. Interrogation of immune-related terms showed downregulation of BCR signaling and upregulation of its inhibitor CD22 specifically in case 8. In contrast, case 9 showed an upregulation of terms associated with infection and immunodeficiency in IDC, including FGGR- and FCER1-mediated signaling.

To further investigate immune-related differences, we calculated ssGSEA for curated immune-related gene signatures (Fig. 3B; Supplementary Methods). Enrichment scores for immune signatures were variable even in DCIS, as seen in the Abba cohort. 8DCIS was “immune hot” and was enriched for macrophage and lymphocytic infiltration signatures, which persisted in IDC despite reduced TILs by H&E. 8IDC showed a higher enrichment for the cytotoxic IFN signature and reduced enrichment for naïve immune response, suggesting an inefficient adaptive immune response in the DCIS. In contrast case 9 had a “colder” immune microenvironment that was maintained in IDC. Immune cell deconvolution based on RNA-sequencing data using a number of methods suggested a reduction of B cells, dendritic cells, and CD4+ T cells in DCIS to IDC (Fig. 3C; Supplementary Fig. S4B).

Further compositional analysis in the Abba cohort showed an enrichment for dendritic cells normal breast tissue, and B cells in DCIS (Fig. 3D). Immune signature analysis presented significant upregulation of macrophage, naïve, lymphocytic, and cytolytic signatures in both ER+.
and ER− DCIS compared with normal. Contrasting ER+ and ER− tumors, we found an enrichment of B cells in ER+ DCIS, whereas ER− IDC was enriched for most immune cell types (Fig. 3E). These results suggest that increased immune infiltration in DCIS may exert both antitumor and immunosuppressive responses.

To obtain a more comprehensive view of immune composition, we performed cycIF, contrasting immune “hot” Case 8 with immune “cold” Case 9 (Fig. 4A; Supplementary Fig. S4B). In contrast to our RNA-data, we observed a decrease in CD8+ accompanied by an increase in immunosuppressive Foxp3+ Treg cells. GZMB+ CD8+ T cells was a rare population that increased from 1.8% of CD8+ T cells in 8DCIS to 8.2% in 8IDC. One population observed specifically in case 8 was the exhausted Ki67+ PD1+ CD4+ T cells, which increased in IDC.

Spatial pattern analysis showed colocalization of the different T-cell subtypes in DCIS, often within tertiary lymphoid structures, rather than interaction with the tumor cells as indicated by low MH indices.
and the low interacting z-scores (Fig. 4D and E), CD8⁺ T cells had a higher interacting score than expected with a random spatial distribution with the tumor in 8DCIS, which was reversed in 8IDC where a higher proportion of Tregs were in close contact with the tumor (Fig. 4D). This observation held true beyond an interacting distance of 10 μm, where CD8⁺ T cells were closer to and Tregs further from tumor cells in DCIS, whereas the reverse was observed in IDC (Supplementary Fig. S4E).

In contrast, case 9 had a low proportion of immune cells, which slightly increased within the proflled IDC region (Fig. 4C). The main immune population present in both DCIS and IDC were CD68⁺ macrophages, which had greater interaction with tumor cells in IDC (Fig. 4A, D and E). An increase in B cells and decrease in Tregs was observed in IDC, and the proportion of CD4⁺ and CD8⁺ (including GZMB⁺) T-cells remained similar within this transition (Fig. 4A), consistent with RNA-sequencing data. In DCIS, both CD8⁺ T cells and
Tregs were in close proximity to the tumor (Fig. 4D) and while not significant by permutation testing at a 10-μm distance, Tregs continued to be closest to tumor cells in IDC (Supplementary Fig. S4F). The low MH index index between CD8+ T cells with any other cell type suggests spatial exclusivity, which in conjunction with its low frequency may result in difficulty in coordinating any cytotoxic response (Fig. 4E).

In both cases, greater intermixing between CD4+ and CD68+ cells with tumor cells was observed in IDC. Colocalization was observed between Tregs with both CD8+ and CD4+ cells; and CD68+ cells with CD4+ cells in DCIS (Fig. 4E) consistent with our H&E analysis, which could be a reflection of a coordinated immune response in DCIS which is diminished in IDC.

Our data shows that the composition and spatial organization of immune cells can be highly variable in DCIS and IDC. One specific "immune hot" TN DCIS becomes more immunosuppressive in IDC through the recruitment of Tregs that intermix with tumor cells. In contrast, an "immune cold" ER+ DCIS shows a population dominated by macrophages and Tregs close to the tumor and has similar expression and compositional properties to IDC. These immune-related differences could be due to differences in tumor subtype and indicate that these differences are already present in the preinvasive stage.

Spatial heterogeneity of tumor cells in the DCIS-to-IDC transition

To assess heterogeneity within tumor cells, we identified 8 major populations based on cyclF (Fig. 4B, Supplementary Fig. S4D). The most dominant cluster had high expression of cytokeratins (CK7,8,17,19) and CDH1; however, subpopulations with decreased cytokeratin expression, particularly of cytokeratin 8, were also detected, particularly in the IDC (Fig. 4B). These CKlow populations often displayed heterogeneous expression of CDH1, consistent with observed genomic loss or mutation in this gene.

Most ducts in 8DCIS were surrounded by an intact CK5+CK14+ myoepithelial cell layer, which was lost in IDC. In contrast, 9DCIS displayed showed thinning of this heterogeneous SMA+ or CK5+CK14+ myoepithelial layer, a common feature of DCIS (Supplementary Fig. S4F; ref. 49). Case 8 was more proliferative than case 9 and had high expression of cytokeratins (CK7,8,17,19) and CDH1; however, subpopulations with decreased expression of CDH1, which was primarily located in the outer regions of tumor-filled ducts. However, despite treatment with tamoxifen, the IDC saw a marked increase in this population (12%–45%) and intermingling with CK+CDH1+ cells. These observations are consistent with reports of high intratumor heterogeneity even in IDC (4, 18).

Some of these observed differences could also be due to tumor subtype, because ER+ tumors (Case 9) are more architecturally organized than TN (Case 8) with tumor cells located in immune excluding nests primarily in ER+ tumors (50).

Genetic aberrations associated with immune changes

Given the changes observed in the immune microenvironment, we investigated whether genomic alterations, specifically in MHC-I presentation or immune suppression (51–53), could contribute to this phenotype. Subclonal losses in PSME3, B2M, TAP1/2, and ERAP1/2 were observed in cases 1, 8, and 9 (Fig. 5A) and in 7 of 29 and 2 of 42 patients in the Abba and Lesurf cohorts, respectively (Supplementary Fig. S5A), which may abrogate MHC-I presentation and lead to immune evasion. Correlation analysis showed an association between copy-number and RNA expression in proteasomal proteins involved in MHC-I presentation but not in immune checkpoint proteins, suggesting that MHC-I loss by CNAs could be one mechanism of immune evasion (Fig. 5B).

PTEN loss is known to have some correlations with immune hot/cold tumors (54). Interestingly, case 2 DCIS has a loss in PTEN (Fig. 1B) and the IDC has frame-shift insertion of PTEN (VAF: 0.67; Fig. 1C) suggesting biallelic loss, and the IDC is more immune cold than the DCIS.

We have previously identified subtype-specific focal amplifications including gain of 17q12 chemokine cluster and STAT2-CD274 at 9p24 as potential mechanisms of immune escape (19). To determine other loci associated with differences in immune microenvironment, we divided the genome into 5-Mb long regions and identified 33 that were enriched for immune-related genes (hypergeometric test, FDR < 0.1, Fig. 5C; Supplementary Fig. S5B; Supplementary Table S6; Supplementary Methods). We then used a combined DCIS cohort (Abba and Recurrence cohorts) and The Cancer Genome Atlas (TCGA) to test for associations between copy-number ratios and immune signature scores (37). Amplification of 1q was negatively associated with immune signaling in both ER+ and ER− IDC but not in DCIS, and contains genes associated with inflammation including the PYHIN, ST10A, and IL10 family of genes, and the CD1 family involved in MHC-like presentation of lipids and glycoproteins (Fig. 5D). Notably, 11q22 (encoding for the caspase family of proteins) and 22q12-13 (encoding for the APOBEC family of proteins and IL2RB) are two regions that are positively associated with immune scores in both DCIS and ER− IDC, and losses in these regions are observed at a frequency of 20% in DCIS. Focusing specifically on ER− IDC, we observed a positive association between macrophage and lymphocyte signatures with amplification of the CCR, CXCL, and IL loci at 3p21, 4q13, and 5q31. Losses in these regions are more frequently observed in ER− breast cancers (55), which was indeed observed in TNBC 8IDC (Fig. 5E). Thus, chromosomal gains and losses could contribute to differences in the immune microenvironment between ER+ and ER− tumors.

Hotspot mutations are predicted to be neoantigens

Cells presenting mutant peptides on MHC-class I can be detected and eliminated by the immune system. For each patient, we determined the HLA types (43) and predicted which oncogenic mutations may present as neoantigens (Supplementary Fig. S5C). As expected, higher mutational burden correlated with higher neoantigen load (p = 0.92, P << 0.001) and neoantigen load correlated with TIL counts in DCIS (p = 0.5, P = 0.006, Recurrence and Abba cohorts).

RNA-expressed neoantigens included CDH1, PIK3CA, which were the same hotspot driver mutations preserved in the DCIS-to-IDC transition, alongside other reported driver events including PTEN, AKT1 (Fig. 6A). Other expressed neoantigens included SF3BI, RAD50, and AKT2. In the Abba cohort we identified PIK3CA (E545K and H1047R), TP53 (R116W), and GATA3 (P95S) mutations as potential neoantigens supported by RNA-sequencing (Fig. 6B).

Given the high frequency of PIK3CA and TP53 mutations predicted as neoantigens in DCIS (14% of patients), we investigated its frequency in the IDC TCGA cohort to determine whether these neoantigens can progress through the DCIS-IDC bottleneck (ref. 37; Fig. 6C; Supplementary Table S7). The proportions of patients with neoantigens between DCIS and IDC were similar, except for GATA3 which saw a reduction, and PIK3CA, which saw an increase in frequency.
Mutations in PIK3CA occurred in 30% of patients and 95% of all PIK3CA mutations were predicted to be neoantigens, predominantly at E545K and H1047R (Supplementary Fig. S5D). Similarly, 60% of TP53 mutations were predicted to be neoantigens, which was much higher than the proportion in other commonly mutated genes including CDH1, GATA3, PTEN, MAP3K1, and KMT2C. Despite TTN and MUC16 being predicted as neoantigens in a high proportion of patients, the mutation sites are variable and do not frequently occur at hotspot or rsSNP sites. PIK3CA and TP53 neoantigens were predicted to bind multiple HLAAs common across the population, including HLA-A:01:01 and HLA-A:02:01 (Fig. 6D; Supplementary Fig. S5E), supporting potential presentation as a neoantigen.

We hypothesized the persistence rather than elimination of neoantigens in the DCIS-to-IDC transition could be due to immunomodulatory effects by tumor cells bearing the mutation. We compared immune signatures of tumors with wild-type versus mutant genes predicted to be neoantigens, accounting for subtype and stage. ER+ patients harboring neoantigenic mutations in PIK3CA, TP53, and CDH1 are associated with increased macrophage signature (Fig. 6E). PIK3CA mutation was also associated with higher TGFβ signaling, and TP53 mutation was associated with IFNγ signaling and wound healing response. Neoantigenic PALB2 was also associated with immune signaling, including IFNγ response.

Our data show that hotspot mutations in common oncogenes including PIK3CA and TP53 can be predicted as neoantigens but are not eliminated in the DCIS-to-IDC transition, and may be associated with immune response. PIK3CA and GATA3 mutations are more common in immune-cold luminal breast tumors (50), whereas TP53 mutations are more common in TIL-rich ER+ tumors. Thus, breast tumor subtype–specific differences in the immune environment might also influence these results.

B-cell receptor repertoire in DCIS and IDC

We assessed the evolutionary trajectory of the microenvironment through profiling of the B-cell receptor (BCR) and TCR repertoire, which can serve as molecular barcodes to monitor changes in subpopulations of T or B cells. We focused on changes in the BCR repertoire due to its higher richness than TCR (Supplementary Fig. S5F), as well as the fact that B-cells play an active role in antitumor
Figure 6.

Neoantigen prediction in DCIS and IDC. A and B, Predicted neoantigens supported by expression of the mutation in RNA-sequencing data in Recurrence cohort (A) and Abba cohort (B). C, Frequency of mutation, neoantigen and rsSNP sites in the most commonly mutated genes in breast cancer in DCIS (Abba and recurrence cohorts) compared with IDC (TCGA cohort). Differences computed using proportion test * P < 0.05. D, HLAs predicted to bind to the most common TP53 and PIK3CA mutant peptides in the TCGA cohort. Asterisked are HLAs predicted to recognize these peptides in the Recurrence/Abba cohort. E, Heatmap showing relative contribution of specific neoantigen to immune signaling pathways in TCGA ER+ patients using a generalized linear model (P < 0.1 shown). F, Changes in BCR repertoire diversity in DCIS and IDC in case 8 and 9. G, Changes in BCR repertoire. Only clonotypes appearing at frequency >1% are shown, and colored clonotypes are shared between samples.
immune responses both by regulating T cells and also by direct antibody production. Using the Shannon Equitability Index as a metric of BCR diversity, we found a reduction in 9IDC compared with the DCIS suggesting a clonal expansion, whereas the reverse was observed in case 9 (Fig. 6F). Indeed, focusing on clones that comprise at least 1% of the BCR repertoire, we found an expansion of the clonotype “CMQRDFFPL7T” in 9IDC, which was also detected at low frequency in the IDC stroma (Fig. 6G).

In case 9, the dominant clonotypes were present in all samples (with the exception of 9DCIS which had lower overall read depth) suggesting little change in the immune microenvironment between DCIS-to-IDC. One dominant clone (“CSSYTSSTLTFV”) detected in 9DCIS was lost in 9IDC, although it was detected at low frequency in the IDC-adjacent tissue. This CDR3 chain can recognize the extracellular domain of HER2-precursor peptides based on the IEDB database (46). This patient is strongly ER and scored as HER2 1+ by IHC.

**Discussion**

In this study, we investigated the genomic profiles of matched pure DCIS and recurrent IDC from an immunologic perspective, which provided an invaluable snapshot of how both the tumor and immune system can collectively influence tumor evolution.

It remains unknown whether recurrent IDC arising many years after diagnosis of pure DCIS is genetically related to the initial lesion. Our study demonstrates this is the case in three of our six cases. Our data suggests that CNAs are early clonal events that are common to both DCIS and IDC, consistent with other studies supporting a linear evolutionary trajectory (3, 4, 12, 56). However, our mutational spectra were vastly different between DCIS and IDC, with many coding mutations lost during this transition. Nevertheless, the presence of conserved mutations in genes including PIK3CA, TP53, CDH1 in matched samples of both DCIS-IDC and normal-DCIS suggest a fitness benefit of these mutations. Studies of matched synchronous cases have also shown DCIS or IDC-specific mutations in select cases and the expansion of a minor subclone in the progression from DCIS to IDC (9, 57), suggesting the possibility of selection but also consistent with models of clonal expansion of a dominant clone under neutral influence tumor evolution.

We have observed a spectrum of immune hot and cold tumors as early as in DCIS, consistent with other immunofluorescence-based studies pure DCIS (19, 20, 58, 59). The increase of TILs in DCIS was shown to be associated with both activated and immunosuppressive gene signatures, suggesting that whilst the immune system may be recruited to the site of the lesion, immunosuppressive pathways may be upregulated to avoid detection. Spatial analysis demonstrated similar metrics for TIL-immune mixing even after accounting for global tissue architecture and composition, highlighting that in our cohort TILs have similar capacities to interact with the tumor front in both DCIS and IDC but higher immune-immune mixing is characteristic of DCIS.

We have noted subtype-specific differences in immune composition, specifically ER tumors are more immune hot compared with ER tumors, in line with previous observations (11, 58, 60, 61), and these differences could be attributed to subtype-specific genetic properties. Not only do TNBC and HER2 tumors have a higher average mutational burden compared with their ER counterparts (62), but are also enriched for CNAs in immune-related regions including the CD274 loci at 9p21 in a subset of TNBC (19) and chemokine cluster at 17q21 in HER2 tumors (19, 63). We additionally show an association between chromosomal loss in chemokine and IL loci with reduced immune signaling in ER tumors (55). TP53 mutation, genomic instability, and telomere crisis are more common in ER tumors and have been associated with higher TILs, which could be recruited following activation of the cGAS-STING pathway mediated by cytosolic double-stranded DNA (47, 60, 61). We found that TP53 neogen-presenting tumors were associated with both activated IFN and inflammatory wound healing macrophage signatures. Indeed, TP53 mutation has been associated with increased CD3+, CD4+, and Foxp3+ cells (61), suggesting that immunosuppressive pathways are also upregulated to counteract this antitumor response. Hence, copy-number aberrations at immune-rich loci may be a mechanism of immune evasion specifically in ER tumors.

We have found that some well-characterized oncogenic mutations in genes including CDH1, and PIK3CA are present in both DCIS and IDC despite being predicted to present as neoantigens. In fact, the PIK3CA E545K mutation, alongside KRAS G12D and BRAF V600E, is one of the most commonly predicted neoantigens in invasive cancer independent of disease type, and studies are underway to determine the potential of these as public neoantigens (37). These data suggest that cells harboring these mutations may upregulate immune-modulatory mechanisms to avoid detection and elimination. While secondary immune effects have not been historically studied, a number of recent studies have begun to dissect the immunologic effect of specific oncogenic transformations. For example, TP53 loss through genome instability has been associated with higher TILs in breast cancer (47, 60, 61). PIK3CA E545K activation in combination with CDH1 loss results in an immunocompromised phenotype with increased macrophage and Treg infiltration in a mouse model of lobular breast cancer (64). Furthermore, PIK3CA mutation is associated with a mesenchymal, secretory phenotype (65), and can create an inflammatory environment through the secretion of cytokines. These studies have highlighted that mutations in driver genes including PIK3CA may directly modulate immune activity to evade immune detection and contribute to the immune-cold phenotype common in ER+ cancers. While functional studies are warranted to validate the immunologic consequences of genetic events, these results present a starting point in evaluating the emergence of subtype-specific differences from an immunogenomics perspective.

The profiling of DCIS has been particularly challenging due to small patient cohorts attributed to long latency periods between DCIS and IDC, low recurrence rates, as well-challenging molecular characterization due to low patient material and artifacts from FFPE preservation. Our cohort is particularly limited in size due to the added challenges in collecting matched cases of pure DCIS, and we have complete data for two cases. Nonetheless, we have showcased an integrative approach to profile this transition using complementary technologies and datasets that have yielded consistent molecular insights. Given that genetic drivers of this transition still remain largely unknown and there are strong relationships between subtype, genomic profiles, and immune microenvironment, integrating data from different cohorts that may have been profiled with different technologies will be essential in uncovering the mechanisms of this transition.

In summary, we profiled matched pure DCIS and recurrent IDC samples to provide an invaluable snapshot of the tumor trajectory in the immune context. This concordant interrogation of both types of information provides a more comprehensive overview into the dynamic interplay between tumor cells and the microenvironment, thereby advancing the search for potential drivers of tumor progression and mechanisms of immune escape.
Trinh et al.

Authors’ Disclosures
A. Trinh reports grants from Helen Gurley Brown Pussycat Foundation during the conduct of the study. S.A. Shukla reports non-financial support from Bristol-Myers Squibb. S.A. Shukla, C. Wu, and J. Gray report grants from National Cancer Institute; grants from The Cancer Society; and personal fees from Convergent Genomics outside the submitted work. J. Jeong reports grants from Komen Foundation, grants from Prospect Creek Foundation, and grants from OHSU Foundation during the conduct of the study. J. Eng reports grants from NIH/NCI during the conduct of the study; grants from Komen Foundation, grants from Prospect Creek Foundation, and grants from SBIR outside the submitted work. C. Wu reports other from BioNTech outside the submitted work. J. Gray reports personal fees from New Leaf Ventures during the conduct of the study; personal fees from PDX Pharmaceuticals, personal fees from KromaTide, personal fees from Quantitative Imaging, personal fees from Health Technologies, and personal fees from Convergent Genomics outside the submitted work; in addition, Dr Gray has a patent for FSH Patent licensed to Abbott Diagnostics. K. Polya reports other from Scorpion Therapeutics, other from Activon Therapeutics, and from t2XAR outside the submitted work. No disclosures were reported by the other authors.

Authors’ Contributions
A. Trinh: Data curation, formal analysis, funding acquisition, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing.
C. Gil Del Alcazar: Investigation. S.A. Shukla: Data curation, software, formal analysis.
K. Polya: Conceptualization, resources, supervision, funding acquisition, writing–original draft, project administration, writing–review and editing.

Acknowledgments
We thank Drs. Deborah Dillon, Paul Spellman, Doris Tabassum, Michalina Janiszewska, and members of the Polya lab for their critical reading of our manuscript. We thank Drs. Thomas O. McDonald, Hua-Jun Wu, and Daniel Temko for their useful suggestions with computational analyses. This work was supported by the National Cancer Institute R35CA197623 (to K. Polya), U01CA195469 (to K. Polya and J. Gray), NIH-U24CA224331 (to C. Wu), R50CA211442 (to S.A. Shukla), U54CA209988 (to K. Polya and J. Gray), DFCI Helen Gurley Brown Presidential Initiative Award (to A. Trinh), and the Breast Cancer Research Foundation (to K. Polya).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 28, 2020; revised November 19, 2020; accepted December 14, 2020; published first December 18, 2020.

References


Molecular Cancer Research

Genomic Alterations during the *In Situ* to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System

Anne Trinh, Carlos R. Gil Del Alcazar, Sachet A. Shukla, et al.


<table>
<thead>
<tr>
<th>Updated version</th>
<th>Access the most recent version of this article at: doi:10.1158/1541-7786.MCR-20-0949</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supplementary Material</td>
<td>Access the most recent supplemental material at: <a href="http://mcr.aacrjournals.org/content/suppl/2020/12/18/1541-7786.MCR-20-0949.DC1">http://mcr.aacrjournals.org/content/suppl/2020/12/18/1541-7786.MCR-20-0949.DC1</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cited articles</th>
<th>This article cites 65 articles, 13 of which you can access for free at: <a href="http://mcr.aacrjournals.org/content/19/4/623.full#ref-list-1">http://mcr.aacrjournals.org/content/19/4/623.full#ref-list-1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Citing articles</td>
<td>This article has been cited by 1 HighWire-hosted articles. Access the articles at: <a href="http://mcr.aacrjournals.org/content/19/4/623.full#related-urls">http://mcr.aacrjournals.org/content/19/4/623.full#related-urls</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-mail alerts</th>
<th>Sign up to receive free email-alerts related to this article or journal.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprints and Subscriptions</td>
<td>To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a>.</td>
</tr>
<tr>
<td>Permissions</td>
<td>To request permission to re-use all or part of this article, use this link <a href="http://mcr.aacrjournals.org/content/19/4/623">http://mcr.aacrjournals.org/content/19/4/623</a>. Click on &quot;Request Permissions&quot; which will take you to the Copyright Clearance Center's (CCC) Rightslink site.</td>
</tr>
</tbody>
</table>