**TITLE**

# Tumor microenvironmental determinants of high-risk DCIS progression

**AUTHORS**

Alexa Glencer[1], Kirithiga Ramalingam[1], Nicole Schindler[1], Hidetoshi Mori[2], Prachi Ghule[4], Kyra Lee[4], Daniela Nachmanson[3], Adam Officer[3], Olivier Harismendy[3], Janet Stein[4], Gary Stein[4], Mark Evans[4], Donald Weaver[4], Christina Yau[1], Gillian L Hirst[1], Michael J Campbell[1], Laura J Esserman[1] and Alexander Borowsky[2]

[1] University of California, San Francisco; [2] University of California, Davis; [3] University of California, San Diego; [4] University of Vermont

## ABSTRACT

Ductal carcinoma *in situ* (DCIS) constitutes an array of morphologically recognized intraductal neoplasms in the mammary ductal tree defined by an increased risk for subsequent invasive carcinomas at or near the site of biopsy detection. However, only 15-45% of untreated DCIS cases progress to invasive cancer, so criteria to predict progression are needed to avoid overtreatment. Furthermore, mechanisms of progression may yield targets for non-surgical therapy and cancer prevention. This study was designed to characterize the tumor microenvironment and molecular profile of high-risk DCIS with predisposition for synchronous or metachronous invasive breast cancer. All patients had DCIS lesions >5cm in size with at least one additional high-risk feature: young age (<45 years), high nuclear grade, hormone receptor negativity, HER2 positivity, the presence of comedonecrosis, or a palpable mass. The tumor immune microenvironment was characterized using multiplex immunofluorescence to identify immune cells and their spatial relationships within the ducts and stroma. Gene copy number analysis and whole exome DNA sequencing identified the mutational burden and driver mutations, and quantitative whole-transcriptome/gene expression analyses were performed. There was no association between the percent of the DCIS genome characterized by copy number variants (CNAs) and recurrence events (DCIS or invasive). Mutations, especially missense mutations, in the breast cancer driver genes *PIK3CA* and *TP53* were common in this high-risk DCIS cohort (47% of evaluated lesions). Tumor infiltrating lymphocyte (TIL) density was higher in DCIS lesions with TP53 mutations ($p=0.0079$) compared to wildtype lesions, but not in lesions with *PIK3CA* mutations ($p=0.44$). Immune infiltrates were negatively associated with hormone receptor status and positively associated with HER2 expression. High levels of CD3+CD8- T cells were associated with good outcomes with respect to any subsequent recurrence (DCIS or invasive cancer), whereas high levels of CD3+Foxp3+ Treg cells were associated with poor outcomes. Spatial proximity analyses of immune cells and tumor cells demonstrated that close proximity of T cells with tumor cells was associated with good outcomes with respect to any recurrence as well as invasive recurrences. Interestingly, we found that myoepithelial continuity (distance between myoepithelial cells surrounding the involved ducts) was significantly lower in DCIS lesions compared to normal tissue ($p=0.0002$) or to atypical ductal hyperplasia ($p=0.011$). Gene set enrichment analysis identified several immune pathways associated with low myoepithelial continuity and a low myoepithelial continuity score was associated with better outcomes, suggesting that gaps in the myoepithelial layer may allow access/interactions between immune infiltrates and tumor cells. Our study demonstrates the immune microenvironment of DCIS, in particular the spatial proximity of tumor cells and T cells, and myoepithelial continuity are important determinants for progression of disease.

## INTRODUCTION

A non-obligate precursor of invasive breast cancer, ductal carcinoma in situ (DCIS), is defined by confinement of proliferating ductal cells within the basement membrane of mammary ducts (1). The routine use of mammography for breast cancer screening over the past thirty years has increased incidence of these lesions from 3% to greater than 25% of all breast cancers detected (2). Standard treatment of DCIS is surgical resection with radiation and endocrine therapy for hormone receptor positive disease; however, DCIS represents a spectrum of biologic risk (3, 4). A uniform approach to the management of DCIS has not led to an overall reduction in invasive breast cancer, which suggests that both overtreatment of indolent lesions and inadequate treatment of more aggressive lesions are contributing factors (5, 6).

Clinical characteristics such as young age and tumor characteristics such as large size, high-grade, hormone receptor (HR) negative status, HER2 positive (HER2+) status, the presence of comedonecrosis, and palpability have each been shown to increase risk of DCIS progression to invasive breast cancer (7-9). Yet despite identification of these risk factors, much remains unknown about the biology of high-risk DCIS.

Some studies have suggested that DCIS lesions with concomitant invasive components have similar genomic profiles to invasive disease (10, 11), but others have demonstrated that genomic alterations occur with the progression of DCIS to invasive cancer (12, 13). While no single immune cell population present in the DCIS tumor immune microenvironment has been found to have independent prognostic value, certain combinations of immune cell infiltrates are associated with high-risk clinical features (14). The proximity of these populations to one another and presence in the ductal versus stromal compartments has potential implications for future breast cancer events (15, 16). The density of tumor infiltrating lymphocytes in the immune microenvironment has also been associated with certain gene expression profiles, which has potential implications for the development of targeted therapeutics (17).

In this study, we aim to characterize the genomic and tumor immune microenvironment characteristics of high-risk DCIS, defined by large size (>5cm) and at least one additional clinicopathologic attribute, and to associate features of the genomic and immune landscape with DCIS recurrence or development of invasive breast cancer. These are high-risk lesions that did not all evolve into invasive breast cancer. Thus, our study is designed to improve our understanding of the biology that distinguishes indolent DCIS from disease with metastatic potential. Without accurate predictors of disease progression, many patients with DCIS undergo surgical resection with no survival benefit, while others with biologically aggressive disease are undertreated and would benefit from targeted molecular and immune therapy. Understanding the underlying genomic profile of high-risk DCIS and the interaction between these lesions and their immune microenvironment is an opportunity to refine prevention strategies and develop more personalized treatment regimens.

## METHODS

### Study population and clinical data collection

For this retrospective study, we identified cases with large DCIS lesions (≥ 5 cm) and at least one of the following high-risk features: young age (<45 years), high nuclear grade, hormone receptor negative, HER2 positive, the presence of comedonecrosis, or a palpable mass. Retrospective clinical characteristics were obtained through chart review without direct patient contact. **Supplementary Figure S1** summarizes the data for the patients in this cohort. All patients enrolled on this study provided written consent for their tissue to be used for research. The Institutional Review Board (IRB) of the University of California, San Francisco (UCSF) approved the conduct of this study with other University of California sites and the University of Vermont (UVM) listed as collaborating sites under protocol #16-20550.

### Single-plex immunohistochemistry

Immunohistochemistry (IHC) was performed as previously reported.(18) Formalin fixed paraffin embedded (FFPE) tissue sections (4 μm) were prepared on Superfrost Plus microscope slides (Thermo Fisher Scientific). Sections were deparaffinized followed by antigen retrieval in CC1 buffer (pH 9, 95°C; Roche), endogenous peroxidase blocking, and then incubation with primary antibodies. Chromogenic detection was performed with VECTASTAIN Elite ABC-HRP kit (Vector Laboratories) followed by counterstaining with hematoxylin. Images were recorded with the Aperio Leica ScanScope XT (Leica Biosystems, Richmond, IL) using x20 objective lens. Primary antibodies used for this study are listed in **Supplementary Table S1**.

### Multiplex immunofluorescence analysis with multispectral imaging

Multiplex immunofluorescence (mIF) analysis was performed with tyramide signal amplification using Opal fluorescent dyes (Akoya Biosciences). Three mIF panels were used in this study (IP1.2, IP2.2, and SP1.1). IP1.2 consisted of antibodies against CD3, CD20, Foxp3, HLA-DR, Ki67, and pancytokeratin AE1/AE3. IP2.2 consisted of antibodies against CD3, CD8, CD68, PD-1, PD-L1, and pancytokeratin AE1/AE3. SP1.1 consisted of antibodies against CD3, CD34, CD68, CD90, αSMA, and p63. Staining was performed with a BondRX autostainer (Leica Biosystems). Antibody/Opal dye pairings, concentrations, and order in the panel were optimized as previously reported (19). Antibody clones, sources, and pairings with Opal dyes are listed in **Supplementary Table S1**.

Slides stained with mIF panels IP1.2 and IP2.2 were scanned using the Vectra Polaris multispectral imaging system (Akoya Biosciences). Regions of interest were selected in Phenochart (Akoya) and spectrally unmixed using inForm software (Akoya). Unmixed image tiles were then stitched together and cell segmentation performed in QuPath (20). Cell phenotyping was performed using flow cytometry based R packages (flowDensity, flowCore, openCyto). Twenty-one cell populations (reported as a percentage of total cell counts), 4 cell:cell ratios, and 2 immune checkpoint scores were identified with these two panels (**Supplementary Table S2**). PD-L1 staining was assessed from IP2.2 using the tumor proportion score (TPS) and the combined positive score (CPS). The TPS was defined as the number of PD-L1$^+$ tumor cells divided by the total number of tumor cells multiplied by 100. The CPS was defined as the number of PD-L1$^+$ tumor cells and PD-L1$^+$ immune cells, divided by the total number of tumor cells multiplied by 100.

Slides stained with mIF panel SP1.1 were scanned on a Nikon A1R HD laser scanning confocal microscope equipped with a spectral detector and 405 nm, 445 nm, 488 nm, 514 nm, 561 nm, 640 nm laser lines (Nikon Instruments Inc., Melville NY). The mIF images obtained on the confocal microscope were spectrally unmixed and analyzed using HALO and HALO AI

analysis packages (Indica Labs). Using a custom trained HALO AI DenseNet classifier, ROIs were classified as either tumor or stroma. The tumor regions were further designated as normal, DCIS or ADH based on a pathologist's annotations on the H&E images. Segmentation and phenotyping of cells within these classified regions was performed with HALO module "HighPlex FL" (version 3.2.1). Seven cell populations were phenotyped with this panel and are reported as cell densities (counts/mm$^2$) in the overall tissue or separately within DCIS or stromal areas (**Supplementary Table S2**).

**Analysis of spatial relationships of cells in the microenvironment of DCIS**

For slides stained with IP1.2 and IP2.2 we employed spatial point pattern analyses using the R package "spatstat" (19). To study the spatial colocalization of cancer cells and immune cells, each mIF image was virtually divided into non-overlapping hexagons (100 μm sides). The number of cancer cells and immune cells within each hexagon was counted. To calculate colocalization of two cell types, A & B, we applied the Morisita-Horn (MH) similarity index (20, 21) to the data using the R package 'divo'. The MH index ranges from 0 (no cells of type A share the same hex with a cell of type B), indicating highly segregated cell populations, to 1 (an equal number of type A and type B cells in each hex), indicating the two cell types are highly colocalized. MH indices were calculated for 23 pairs of cell types (**Supplemental Table S2**).

To further examine spatial relationships, we utilized the nearest neighbor distribution function, *G(r),* to quantify the relative proximity of any two cell types. This function is a spatial distance distribution metric that represents the probability of finding at least one cell of a given type within a radius (*r*) of another cell type. The *G(r)* function computations were performed using the R package "spatstat". We defined a Spatial Proximity Score (*SPS*) as the area under the *G(r)* curve (AUC) for $r$ = 0 to 20 μm. A large *SPS* indicates close proximity between the cell populations analyzed. The *SPS* is directional. For example, the score for T cells within 20 μm of tumor cells, denoted *Tum_T.SPS* will be different than the score for tumor cells within 20 μm of T cells (*T_Tum.SPS)*. *SPS* scores were determined for 44 pairs of cell types (**Supplemental Table S2**). The 20 μm interval was selected to build upon prior work in the field (22-24). To put this into perspective, a lymphocyte has a diameter of ~10 μm.

Spatial analyses from slides stained with SP1.1 were performed with HALO Highplex FL (version 3.2.1) using the Proximity Analysis algorithm from the Spatial Analysis module within HALO. The distance between cells positive for each marker and cells that were co-positive for p63 and αSMA was measured and is reported as "marker" DistFromEpithelium. In order to quantify myoepithelial continuity, the average nearest neighbor distances between $\alpha SMA^+/p63^+$ myoepithelial cells within the regions of DCIS were calculated. A myoepithelial continuity score was then derived by reverse scoring these distances such that a high score (high continuity) = short distance between myoepithelial cells (close proximity) and a low score (low continuity) = greater distance between these cells. Reverse scoring was performed by taking the maximum average nearest neighbor distance between myoepithelial cells, adding 1 to this distance, and then subtracting the nearest neighbor distances from this value to get the reverse scored values.

**Gene expression**

FFPE tissue sections were sent to Agendia, Inc, for RNA extraction and gene expression on Agilent 32K expression arrays (Agendia32627_DPv1.14_SCFGplus with annotation GPL20078). MammaPrint risk classification (High vs. Low risk) and BluePrint molecular subtype (Luminal, HER2 or Basal type) were calculated. For each array, the green channel mean signal of each probeset was log2-tranformed and centered within array to its 75$^{th}$ quantile as per the manufacturer's data processing recommendations. All values flagged for

non-conformity are NA'd out; and a fixed value of 9.5 was added to avoid negative values. Probeset level expression data was mean-collapsed to gene level by gene symbol, yielding a normalized expression dataset comprising 54 samples from 52 patients x 22584 unique genes. Twenty-eight previously published immune-related signatures, along with a cell proliferation and ER/PR signature, were computed using the scoring methods described in Supplemental Table 3 from the gene-level normalized expression data.

Gene set enrichment analysis (GSEA) was used to determine pathways enriched in DCIS with high or low myoepithelial continuity scores. GSEA was performed using the R package fgsea (v1.24.0) (25) and the human MSigDB collection of gene sets derived from the GO biological process ontology (c5.go.bp.v2023.1.Hs.symbols.gmt). Pathways were considered enriched if adjusted p-values < 0.05.

**Whole exome and whole genome sequencing**

Whole exome sequencing including copy number alterations (CNAs) and mutations was performed on 17 cases, and low-pass whole genome sequencing including CNAs was performed on 34 cases. Tissue from DCIS involved ducts was obtained by laser capture microdissection (ArcturusXM system; ThermoFisher). DNA was extracted using the QIAmp DNA Micro Kit (Qiagen). Libraries were prepared using Ultra II FS library preparation kit (New England Biolabs) and amplified with KAPA HiFi HotStart Real-time PCR using KAPA P5 and P7 primers (KAPA Biosystems).

*Exome sequencing.* Exome sequencing was performed as previously described (37). Sequencing libraries were deconvoluted using bcl2fastq. Sequencing data were analyzed using bcbio-nextgen (v1.1.6) as a workflow manager. Adapter sequences were trimmed using atropos (v1.1.22), the trimmed reads were subsequently aligned with bwa-mem (v0.7.17) to reference genome hg19, then PCR duplicates were removed using biobambam2 (v2.0.87). Additional BAM file manipulation and collection of QC metrics were performed with picard (v2.20.4) and samtools (v1.9). Contamination was measured using verifyBamID2 (v1.0.6). Samples with less than 90% of their bases covered at 10x or contamination estimated to be >5% were not used for mutational analysis.

*Low-coverage whole-genome sequencing.* For low-coverage whole-genome sequencing, sample libraries were sequenced to a target coverage of 0.5x using paired-end 2x100bp reads on a NovaSeq 6000 (Illumina). Sequencing libraries were deconvoluted using bcl2fastq. Adapter sequences were trimmed from the raw fastq files using atropos (v1.1.31). The trimmed reads were then aligned to GRCh38 using bwa-mem (v0.7.17). Duplicate reads were then marked using biobambam (v2.0.87). Overall genome-wide coverage was measured using mosdepth (v0.2.6), and contamination was measured using verifyBamID2 (v1.0.6). Samples with less than 0.45x coverage or were estimated to be >5% contaminated were removed from downstream analyses.

*Copy number alterations.* CNAs were called using CNVkit (v0.9.9) in “wgs” mode with average bin size set at 100,000 bp. A set of unrelated normal tissues sequenced with the same protocol were used to generate a panel of normals used during CNA calling. Any bins with a $\log_2$ copy ratio lower than -15 were considered artifacts and removed. Breakpoints between copy number segments were determined using the circular binary segmentation algorithm ($p < 10^{-4}$). Copy number genomic burden was computed as the sum of sizes of segments in a gain ($\log_2(\text{ratio}) > 0.3$) or loss ($\log_2(\text{ratio}) < -0.3$) over the sum of the sizes of all segments. Gene-level CNA and exome-sequencing-based CNA were measured as previously described (37).

*Pathogenic germline variants.* For all cases, we identified variants that were either “Pathogenic” or “Likely-Pathogenic” in ClinVar and had a variant allele frequency (VAF) of at least 15% in the following breast cancer susceptibility genes: *ATM, BARD1, BLM, BRCA1,*

*BRCA2, BRIP1, CDH1, CHEK2, EPCAM, FAM175A, FANC, MLH1, MRE11, MSH2, MSH6, MUTYH, NBN, NBS1, NF1, PALB2, PMS2, PTEN, RAD50, RAD51C, RAD51D, RECQL, STK11, TP53, XRCC, MEN1,* or *PPM1D.*

*Somatic mutations.* Somatic mutations and removal of putative germline variants were determined from tumor-only data as previously described (37).

**Statistical analyses**

Statistical analyses were performed using R. Associations of clinical characteristics with immune cell populations, gene expression signatures, and spatial metrics are presented as dot matrix plots. Associations of immune biomarkers with chromosomal gains/losses are also presented as dot matrix plots. Grouped data box plots are presented with individual sample points. Data were compared between two groups with Wilcoxon's test or between multiple groups using the Kruskal-Wallis test. Associations of biomarkers with subsequent recurrence events were assessed in univariate analyses using Kaplan-Meier curves and the Log Rank test statistic. Multiple testing corrections were not performed.

## RESULTS

### Study population

For this retrospective study, we identified cases with large DCIS lesions (≥ 5 cm) and at least one of the following high-risk features: young age (<45 years), high nuclear grade, hormone receptor (HR) negative, HER2 positive, the presence of comedonecrosis, or a palpable mass. In total, 61 cases were identified for this study. **Supplementary Figure S1** provides an overview of the clinicopathologic characteristics of this cohort. The median age at diagnosis was 45 years (range: 30-88) with a median follow-up time of 8 years. The majority of patients (64%) were diagnosed with high-grade DCIS, and most cases were associated with comedonecrosis (77%). Sixty-two percent of cases were hormone receptor (HR) positive and 57% were HER2+. Lesions were subtyped as HR+HER2+ (n=21; 34%), HR+HER2- (n=17; 28%), HR-HER2+ (n=14; 23%) or triple negative (n=9; 15%). There were 14 recurrence events in this cohort; 6 DCIS and 8 invasive breast cancers.

### Biomarker assay platforms

We utilized three assay platforms to characterize the DCIS lesions and their immune microenvironments (**Figure 1**). Agilent 44K expression array data were generated from 85% (n=52) of the 61 cases. Twenty-eight gene expression signatures identifying immune cell populations, immune checkpoint genes, or immune signaling pathways were evaluated (**Supplementary Table S3**). Genomic copy number data and whole exome sequencing data were generated from 56% (n=34) and 29% (n=17) of the cases, respectively. Staining FFPE tissue sections with single-plex IHC markers and mIF panels IP1.2, IP2.2, and SP1.1 was successfully performed on 89% (n=54), 87% (n=53), 93% (n=57), and 97% (n=59) of cases, respectively (**Figure 1**). Fifty-two biomarkers (immune or tumor cell populations, immune: immune cell ratios, immune checkpoint scores) were identified and evaluated from these staining panels (**Supplementary Table S2**).

Several biomarkers were identified by both mIF and gene signatures. We evaluated correlations of these biomarkers across platforms (**Supplementary Figure S2**) and found strong positive correlations between percent TILs (by mIF) and the TIL ($p = 1.6\text{x}10^{-6}$) and Module4_TcellBcell_score ($p = 1.5\text{x}10^{-11}$) expression signatures. Percentages of T cells and B cells were also strongly positively correlated with T_cell and B_cell expression signatures ($p = 1.5\text{x}10^{-4}$ and $p = 3.1\text{x}10^{-11}$, respectively), as were $CD8^+$ T cells (CD8pT by mIF) with the CD8_T_cell and Cytotoxic_cell signatures ($p = 1.1\text{x}10^{-4}$ and $p = 4\text{x}10^{-4}$, respectively). In contrast, Treg cell and macrophage percentages did not correlate with their respective gene signatures. Finally, a proliferation gene signature (Module11_Prolif_score) was positively correlated with $Ki67^+$ tumor cells as measured by mIF ($p = 3.1\text{x}10^{-6}$).

### The genomic landscape of DCIS and immune infiltrates

We performed whole exome sequencing (WES) to identify copy number alterations (CNAs) and mutational profiles and evaluate their associations with immune infiltrates. As shown in **Figure 2A**, frequently observed copy number changes included 1q, 8q, 16p, and 17q gains and 8p, 16q, and 17p losses, consistent with observations in other DCIS datasets (26-28). The fraction of the genome altered was not significantly different between the non-recurrent and recurrent cases (DCIS or invasive recurrences; **Figure 2B**). In addition, there was no association with the fraction of the genome altered (overall CNAs, gains only, or losses only) with any immune biomarker (**Supplementary Figure S3**). However, we did observe copy number alterations at several chromosomal regions associated with various immune cell populations (**Figure 2C and Supplementary Figure S3**). In particular, gains on chromosomes 6p, 6q, and 21q, and losses on chromosomes 8p, 9p, and 17p were associated

with higher TIL, T cell, $CD8^+$ T cell, $CD8^-$ T cell, Treg, and/or B cell infiltrates. A cytotoxic cell gene signature and two natural killer (NK) cell gene signatures were also associated with 8p chromosomal losses.

Focusing specifically on breast cancer-related driver genes, the most commonly altered gene with respect to copy number status was the *ERBB2* gene (53% of cases had *ERBB2* gains) (**Supplementary Figure S4**). Other common gains (in over 40% of cases) included the *MYC*, *PPM1D*, and *NBN* loci. Common losses (in over 40% of cases) were observed for *MAP2K4*, *NCOR1*, *TP53*, *GPS2*, and *FAM86B1*. Losses were also noted in genes associated with DNA repair/genome instability (*BRCA1*, *BRCA2*, *TP53, ATRX, ZMYM3*) and cell-cycle regulation (*RB1*). Interestingly, copy number loss of *ATRX* and *ZMYM3*, genes that protect against chromosomal instability, and copy number loss of *CDKN2A*, a tumor suppressor/negative regulator in the CDK4/Rb pathway, were associated with higher densities of TILs, T cells, $CD8^+$ T cells, $CD8^-$ T cells, Treg cells, and/or B cells (**Supplementary Figure S5**).

The mutational status of driver genes commonly altered in breast cancer is illustrated in **Figure 2D** for 17 of the DCIS cases. *PIK3CA* and *TP53* were the most frequently mutated genes. Mutations in the *TP53* gene were significantly associated with high TIL infiltrates (**Figure 2E**) and other immune cell populations (**Supplementary Figure S6**). In contrast, mutations in *PIK3CA* were not associated with any immune cell populations (**Figure 2F** and **Supplementary Figure S6**).

**Immune cell populations are associated with DCIS clinical characteristics and outcomes**

Single-plex IHC and mIF panels IP1.2, IP2.2, and SP1.1 identified 46 immune or tumor cell populations, 4 immune:immune cell ratios, and 2 immune checkpoint scores (**Supplementary Table 2**). Immune cell densities varied across tumors. The fraction of TILs ($CD3^+$ T cells plus $CD20^+$ B cells) ranged from 3 to 59% (median 18%). As shown in **Figure 3A**, several immune cell populations identified by mIF were significantly associated with HR negativity and/or HER2 positivity. These included TILs, T cells, $CD8^+$ T cells, Treg cells, and B cells. Proliferating cells ($Ki67^+$ T cells, B cells, Tregs) and HLADR positive T and B cells were negatively associated with HR status and positively associated with HER2 status. Several PD-1/PD-L1 populations were also associated with HR negativity ($PD\text{-}L1^+$ immune cells, CPS score) and/or HER2 positivity ($PD\text{-}1^+$ T cells, $PD1^+CD8^+$ T cells, $PD\text{-}L1^+$ immune cells, CPS score). The presence of comedonecrosis was positively associated with $PD1^+CD8^+$ T cells, any $PD\text{-}L1^+$ cell (tumor or immune), $PD\text{-}L1^+$ immune cells, and the higher Treg:T cell ratio, and negatively associated with the overall density (cells/$mm^2$) of $CD68^+$ cells and the density of $CD68^+$ cells within the DCIS containing ducts. Palpable lesions were positively associated with the overall density (cells/$mm^2$) of $CD68^+$ cells, $p63^+$ cells, and $CD34^+CD68^+$ cells as well as with the density of $CD34^+$ cells and $CD68^+$ cells in the stroma. Finally, from the single-plex IHC markers analyzed, $BAG1^+$ tumor cells were positively associated with HR status and the IHC4 score (ER/PR/HER2/Ki67) was positively associated with grade and HER2 status, and negatively associated with HR status (**Figure 3A**).

We also evaluated 28 gene expression signatures identifying immune cell populations, immune checkpoint genes, or immune signaling pathways (**Supplemental Table S3**). As with the staining results, many of these expression signatures were associated with HR negativity and HER2 positivity (**Figure 3B**). Unlike TIL and $PD\text{-}1^+$ cell percentages based on mIF staining, the TIL gene signature and PD-1 gene expression were positively associated with grade.

Of the 80 mIF, single-plex IHC, and gene expression biomarkers examined, only 3 were associated with any recurrence event. High numbers of $CD3^+CD8^-$ T cells were associated with good outcomes ($p = 0.027$; **Figure 3C**) while high expression of a Treg gene signature

and high expression of a proliferation gene signature were associated with poor outcomes (p = 0.016; **Figure 3D** and p = 0.0069; **Figure 3E**, respectively). When analysis was restricted to invasive recurrences, only the proliferation gene signature was associated with poor outcomes (p = 0.0042; **Figure 3F**).

**Spatial relationships of cells in the microenvironment of DCIS are associated with clinical characteristics and outcomes**

We investigated associations between clinical characteristics and outcomes with the proximity of various immune cell populations to tumor cells or other immune cells. We utilized the Morisita-Horn index (MH) as a measure of the degree of colocalization of two cell types. When two cell types are highly colocalized, the Morisita-Horn index (which ranges from 0 to 1) will be near 1. In contrast, highly segregated cell populations will generate a Morisita-Horn index value close to 0. The MH index was calculated for 23 pairs of cell types (**Supplemental Table S2**). As shown in **Figure 4A**, 13 of these indices were significantly associated with at least one clinical characteristic. Most of these were associated with HR negativity. The colocalizations of TIL with Treg cells (TIL_Treg.MHI), Treg cells with other T cells (Treg_Foxp3nT.MHI), and macrophages with $CD3^+CD8^-$ T cells (Mac_CD8nT.MHI) were associated with $HER2^+$ cases. Colocalization of T cells with B cells (T_B.MHI), $CD3^+CD8^+$ T cells with $CD3^+CD8^-$ T cells (CD8pT_CD8nT.MHI), and Treg cells with other T cells (Treg_Foxp3nT.MHI) were all positively associated with high grade DCIS.

To further characterize the spatial relationships between cells in the tumor microenvironment, we utilized a nearest neighbor distribution function, *G(r),* to evaluate the probability of a cell of type "*a*" having at least one cell of type "*b*" within a distance *r*. We defined a Spatial Proximity Score (*SPS*) as the area under the *G(r)* function curve, from 0-20μm. *SPS*s were calculated for 44 cell pairs (**Supplemental Table S2**). Twenty-four of these scores were significantly associated with at least one clinical characteristic of high-risk DCIS (**Figure 4A**). Interestingly, various *SPS*s involving PD-$1^+$ cells in proximity to PD-$L1^+$ cells were negatively associated with HER2 expression whereas various other cell type interactions were associated with HER2 positivity.

We also calculated the average distances of $\alpha SMA^+p63^-$, $CD3^+$, $CD34^+$, or $CD68^+$ cells to the nearest $\alpha SMA^+p63^+$ myoepithelial cell. Three of these markers (αSMA_DistFromEpithelium, CD3_DistFromEpithelium, CD34_DistFromEpithelium) were associated with HR negativity (**Figure 4A**).

Of the 71 spatial metrics evaluated, only the Tum_T spatial proximity score was associated with outcomes. A high Tum_T.SPS, indicating close proximity of tumor cells and T cells, was significantly associated with good outcomes for any recurrence (p = 0.047; **Figure 4B**) as well as invasive recurrences (p = 0.026; **Figure 4C**). We next combined the Tum_T.SPS with the proliferation gene signature (Module11_Prolif_score; **Figure 2E & F**) and compared outcomes of patients with a low Tum_T.SPS and high proliferation score (n=13) to all other cases (n=36). Low spatial proximity of tumor cells and T cells combined with a high proliferation score was significantly associated with poor outcomes for all recurrences (p = 0.00027; **Figure 4D**) as well as invasive recurrences (p = 0.00021; **Figure 4E**).

**Myoepithelial continuity is associated with subsequent recurrence events**

Using the nearest neighbor distances between $\alpha SMA^+p63^+$ myoepithelial cells, we derived a myoepithelial continuity score. Myoepithelial continuity was significantly lower in DCIS compared to benign (normal or ADH) lesions (**Figure 5A-G**). Interestingly, low myoepithelial continuity was significantly associated with better outcomes for any recurrence (p = 0.038; **Figure 5H**). This suggests that low myoepithelial continuity may serve a protective function by allowing greater access of immune cells to the DCIS tumor cells. We examined correlations of

various immune cell populations identified by mIF staining or gene expression signatures with the MyoEpiContinuity score and did not find any significant correlations (**Supplementary Figure S7**). We next used gene set enrichment analysis to identify pathways/ontologies that were correlated with low or high myoepithelial continuity. Low MyoEpiContinuity scores were enriched for several immune related ontologies including adaptive immune response, leukocyte proliferation, TNF superfamily cytokine production, and MHC Class II antigen assembly, processing, and presentation (**Supplementary Figure S8**). In contrast, high MyoEpiContinuity score was only enriched for one ontology related to cell adhesion. Finally, given that the presence of $CD3^{+}CD8^{-}$ T cells was associated with good outcomes for any recurrence (**Figure 3C**), we compared outcomes of patients with a low MyoEpiContinuity score plus high $CD3^{+}CD8^{-}$ T cell infiltrates (n=12) to all other cases (n=28). As shown in **Figure 5I**, a Low/High MyoEpiContinuity/CD8nT score was significantly associated with good outcomes for all recurrences ($p = 0.0039$).

**DISCUSSION**

This study was designed to identify genomic alterations, expression signatures, and interactions among the tumor microenvironment, immune cells, stromal elements, and the DCIS epithelium in large (>5cm) clinically high-risk DCIS lesions. Our cohort of 61 cases with long-term follow up (median 8 years) were all treated with standard of care (surgery +/- radiation). 14 of 61 (23%) had a subsequent disease recurrence event, 8 with invasive carcinoma, and 6 with recurrent DCIS. We performed an array of tissue-based assays and microdissections accompanied by expert (ADB) annotation of the areas of DCIS, and also a whole tissue section (bulk) transcriptome analysis. Cross-platform correlations provide new insight into the underlying biologic interactions that appear to constrain or permit progression and may reflect patterns of immunosurveillance/recognition of these lesions.

CNAs and mutations of breast cancer driver genes, most notably TP53, were found to be significantly associated with dense immune cell infiltrates. This is similar to previously reported observations in DCIS and invasive breast cancer (13, 38). Frequently observed CNAs in our cohort included gains on chromosomes 1q, 8q, 16p, and 17q and losses on chromosomes 8p, 16q, and 17p. Breast cancer is considered to be a copy number class tumor, and these CNAs in DCIS lesions are consistent with previously published reports (26-29). While we did not find that the fraction of the DCIS genome altered correlated with immune infiltrates or recurrence events, CNAs at certain chromosomal regions were associated with various immune cell populations. Gains on chromosomes 6p, 6q, and 21q or losses on chromosomes 8p, 9p, and 17p were associated with higher TIL, T cell, $CD3^+CD8^+$ T cell, $CD3^+CD8^-$ T cell, Treg, or B cell infiltrates. ERBB2 was the most prevalent gene with CNAs (53% of cases), which is consistent with a recent analysis of CNAs in DCIS. However, in both this study and our own, CNA at ERBB2 was not associated with recurrence events (29). In our cohort, copy number losses were common in various known breast cancer driver genes, including TP53 (>40% of cases), BRCA1, BRCA2, ATRX, and ZMYM3. Loss of TP53 function leads to overall genomic instability and resistance to DNA damage-mediated apoptosis. ATRX functions to regulate the DNA damage response and ameliorates chromosomal instability (30). ZMYM3 deficiency results in impaired homologous recombination repair and genome instability (31). We observed higher immune infiltrates (TILs, T cells, $CD8^+$ T cells, Treg, and B cells) in DCIS harboring TP53 mutations or losses of ATRX or ZMYM3, which is consistent with the concept that genomic instability may lead to the generation of neoantigens that can activate the immune system (40).

In analyzing the presence, density, and spatial relationships of various immune cells within the tumor immune microenvironment, we identified significant associations with certain clinical characteristics and outcomes. Like their invasive counterparts, we found that HR negative and/or $HER2^+$ DCIS was associated with high densities of TILs, T cells, $CD3^+CD8+$ T cells, $CD3^+CD8^-$ T cells, Treg cells, and B cells as well as proliferating ($Ki67^+$) subgroups of T cells, Treg, and B cells. Several PD-1/PD-L1 immune cell populations were also associated with HR negative or $HER2^+$ DCIS. Gene expression signatures identifying immune cell populations or signaling pathways were associated with HR negative or $HER2^+$ DCIS. These associations are consistent with several published studies (14, 32-36); however, a recent study by Almekinders et al. that examined the relationship of DCIS immune infiltrates with recurrence events failed to find an association (32). It should be noted that the cohort in the Almekinders study differed meaningfully from our own, in that the size of >60% of the cases in their cohort was unknown compared to the requirement that our cases all be at least 5cm in size. Furthermore, more than half of our cases (57%) were $HER2^+$ compared to <30% of their cases (32). Our cohort represents a higher risk population with overrepresentation of intrinsic subtypes known to possess a more immunologically active tumor microenvironment.

In examining 80 possible biomarkers derived from mIF, single-plex IHC, and gene expression assays, we were able to identify three biomarkers that were associated with recurrence events. We found that a high density of $CD3^{+}CD8^{-}$ T cells was significantly associated with a decrease in any recurrence event. High expression of a Treg gene signature and high expression of a proliferation gene signature were each associated with greater risk of overall recurrence; the proliferation gene signature was also associated with greater risk of invasive recurrence specifically. Similar associations of a Treg gene signature and higher proportion of Treg cells in the tumor immune infiltrate with increased recurrence has been observed in the setting of early stage (Ia and Ib) non-small cell lung cancer and non-metastatic renal cell carcinoma (41, 42). Furthermore, we found that close proximity of T cells to tumor cells was significantly associated with better outcomes, with respect to any recurrence or invasive recurrences. These findings suggest that it is not solely the presence of T cells in the tumor immune microenvironment but their spatial relationship to tumor cells that predicts clinical outcomes.

Moreover, the stromal compartment plays a critical immunoregulatory role. We used spatial proximity of $\alpha SMA^{+}p63^{+}$ myoepithelial cells to calculate a myoepithelial continuity score. Consistent with recently published work (16), we found that myoepithelial continuity is significantly lower in DCIS compared to benign (normal or ADH) lesions and that higher myoepithelial continuity among DCIS lesions is significantly associated with a higher recurrence risk. In addition, low myoepithelial continuity was associated with a variety of immune related pathways. These findings raise an intriguing hypothesis that myoepithelial cells may limit host immune recognition of the DCIS cells harboring neoantigens. Indeed, HER2 amplification/overexpression itself may be one of these neoantigens-- with high prevalence in our cohort, but also high prevalence in all DCIS compared to lower prevalence in invasive carcinomas.

In summary, this study demonstrated that DCIS can be stratified based upon features of the tumor immune microenvironment which in turn correlates with intrinsic epithelial subtypes including propensity for high neo-antigenicity, along with myoepithelial continuity and finally immune activation. This has the potential to guide clinical decision-making including the safety of active surveillance (watchful waiting) for lower risk lesions as well as improved therapy including possible immunomodulation in higher risk lesions.

.

## REFERENCES


1. Siziopikou KP. Ductal carcinoma in situ of the breast: current concepts and future directions. Arch Pathol Lab Med. 2013;137(4):462-6.
2. Esserman LJ, Thompson IM, Reid B, Nelson P, Ransohoff DF, Welch HG, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. Lancet Oncol. 2014;15(6):e234-42.
3. Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. Cancer. 2005;103(9):1778-84.
4. Stuart KE, Houssami N, Taylor R, Hayen A, Boyages J. Long-term outcomes of ductal carcinoma in situ of the breast: a systematic review, meta-analysis and meta-regression analysis. BMC Cancer. 2015;15:890.
5. Giannakeas V, Sopik V, Narod SA. Association of a Diagnosis of Ductal Carcinoma In Situ With Death From Breast Cancer. JAMA Netw Open. 2020;3(9):e2017124.
6. Welch HG, Woloshin S, Schwartz LM. The sea of uncertainty surrounding ductal carcinoma in situ--the price of screening mammography. J Natl Cancer Inst. 2008;100(4):228-9.
7. Badve SS, Gökmen-Polar Y. Ductal carcinoma in situ of breast: update 2019. Pathology. 2019;51(6):563-9.
8. Rakovitch E, Nofech-Mozes S, Hanna W, Baehner FL, Saskin R, Butler SM, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. Breast Cancer Res Treat. 2015;152(2):389-98.
9. Williams KE, Barnes NLP, Cramer A, Johnson R, Cheema K, Morris J, et al. Molecular phenotypes of DCIS predict overall and invasive recurrence. Ann Oncol. 2015;26(5):1019-25.
10. Iakovlev VV, Arneson NC, Wong V, Wang C, Leung S, Iakovleva G, et al. Genomic differences between pure ductal carcinoma in situ of the breast and that associated with invasive disease: a calibrated aCGH study. Clin Cancer Res. 2008;14(14):4446-54.
11. Castro NP, Osório CA, Torres C, Bastos EP, Mourão-Neto M, Soares FA, et al. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. Breast Cancer Res. 2008;10(5):R87.
12. Casasent AK, Schalck A, Gao R, Sei E, Long A, Pangburn W, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. Cell. 2018;172(1-2):205-17.e12.
13. Kim SY, Jung SH, Kim MS, Baek IP, Lee SH, Kim TM, et al. Genomic differences between pure ductal carcinoma in situ and synchronous ductal carcinoma in situ with invasive breast cancer. Oncotarget. 2015;6(10):7597-607.
14. Campbell MJ, Baehner F, O'Meara T, Ojukwu E, Han B, Mukhtar R, et al. Characterizing the immune microenvironment in high-risk ductal carcinoma in situ of the breast. Breast Cancer Res Treat. 2017;161(1):17-28.
15. Sharma M, Beck AH, Webster JA, Espinosa I, Montgomery K, Varma S, et al. Analysis of stromal signatures in the tumor microenvironment of ductal carcinoma in situ. Breast Cancer Res Treat. 2010;123(2):397-404.
16. Risom T, Glass DR, Averbukh I, Liu CC, Baranski A, Kagel A, et al. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. Cell. 2022;185(2):299-310.e18.
17. Agahozo MC, Smid M, van Marion R, Hammerl D, van den Bosch TPP, Timmermans MAM, et al. Transcriptomic Properties of HER2+ Ductal Carcinoma In Situ of the Breast Associate with Absence of Immune Cells. Biology (Basel). 2021;10(8).

18. Mori H, Soonsawad P, Schuetter L, Chen Q, Hubbard NE, Cardiff RD, et al. Introduction of Zinc-salt Fixation for Effective Detection of Immune Cell-related Markers by Immunohistochemistry. Toxicol Pathol. 2015;43(6):883-9.
19. Baddeley A, Rubak E, Turner R. Spatial Point Patterns: Methodology and Applications with R: Chapman and Hall/CRC Press; 2015.
20. Horn HS. Measurement of "Overlap" in Comparative Ecological Studies. The American Naturalist. 1966;100(914):419-24.
21. Wolda H. Similarity indices, sample size and diversity. Oecologia. 1981;50(3):296-302.
22. Barua S, Solis L, Parra ER, Uraoka N, Jiang M, Wang H, et al. A Functional Spatial Analysis Platform for Discovery of Immunological Interactions Predictive of Low-Grade to High-Grade Transition of Pancreatic Intraductal Papillary Mucinous Neoplasms. Cancer Inform. 2018;17:1176935118782880.
23. Carstens JL, Correa de Sampaio P, Yang D, Barua S, Wang H, Rao A, et al. Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. Nat Commun. 2017;8:15095.
24. Giraldo NA, Sanchez-Salas R, Peske JD, Vano Y, Becht E, Petitprez F, et al. The clinical role of the TME in solid cancer. Br J Cancer. 2019;120(1):45-53.
25. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. bioRxiv. 2021:060012.
26. Trinh A, Gil Del Alcazar CR, Shukla SA, Chin K, Chang YH, Thibault G, et al. Genomic Alterations during the In Situ to Invasive Ductal Breast Carcinoma Transition Shaped by the Immune System. Mol Cancer Res. 2021;19(4):623-35.
27. Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E, et al. A Molecular Portrait of High-Grade Ductal Carcinoma In Situ. Cancer Res. 2015;75(18):3980-90.
28. Lesurf R, Aure MR, Mørk HH, Vitelli V, Lundgren S, Børresen-Dale AL, et al. Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer. Cell Rep. 2016;16(4):1166-79.
29. Gorringe KL, Hunter SM, Pang JM, Opeskin K, Hill P, Rowley SM, et al. Copy number analysis of ductal carcinoma in situ with and without recurrence. Mod Pathol. 2015;28(9):1174-84.
30. Aguilera P, López-Contreras AJ. ATRX, a guardian of chromatin. Trends Genet. 2023;39(6):505-19.
31. Leung JW, Makharashvili N, Agarwal P, Chiu LY, Pourpre R, Cammarata MB, et al. ZMYM3 regulates BRCA1 localization at damaged chromatin to promote DNA repair. Genes Dev. 2017;31(3):260-74.
32. Almekinders MM, Bismeijer T, Kumar T, Yang F, Thijssen B, van der Linden R, et al. Comprehensive multiplexed immune profiling of the ductal carcinoma in situ immune microenvironment regarding subsequent ipsilateral invasive breast cancer risk. Br J Cancer. 2022;127(7):1201-13.
33. Pruneri G, Lazzeroni M, Bagnardi V, Tiburzio GB, Rotmensz N, DeCensi A, et al. The prevalence and clinical relevance of tumor-infiltrating lymphocytes (TILs) in ductal carcinoma in situ of the breast. Ann Oncol. 2017;28(2):321-8.
34. Agahozo MC, van Bockstal MR, Groenendijk FH, van den Bosch TPP, Westenend PJ, van Deurzen CHM. Ductal carcinoma in situ of the breast: immune cell composition according to subtype. Mod Pathol. 2020;33(2):196-205.
35. Toss MS, Abidi A, Lesche D, Joseph C, Mahale S, Saunders H, et al. The prognostic significance of immune microenvironment in breast ductal carcinoma in situ. Br J Cancer. 2020;122(10):1496-506.

36. Hendry S, Pang JB, Byrne DJ, Lakhani SR, Cummings MC, Campbell IG, et al. Relationship of the Breast Ductal Carcinoma In Situ Immune Microenvironment with Clinicopathological and Genetic Features. Clin Cancer Res. 2017;23(17):5210-7.
37. Nachmanson, D., Officer, A., Mori, H. *et al.* The breast pre-cancer atlas illustrates the molecular and micro-environmental diversity of ductal carcinoma in situ. *npj Breast Cancer* 2022; **8**(6).
38. Zhang Z, Hao R, Guo Q, Zhang S, Wang X. TP53 Mutation Infers a Poor Prognosis and Is Correlated to Immunocytes Infiltration in Breast Cancer. *Front Cell Dev Biol*. 2021;9:759154.
39. Niwińska A, Olszewski WP. The role of stromal immune microenvironment in the progression of ductal carcinoma in situ (DCIS) to invasive breast cancer. *Breast Cancer Res*. 2021;23(1):118.
40. Mardis ER. Neoantigens and genome instability: impact on immunogenomic phenotypes and immunotherapy response. *Genome Med*. 2019;11(1):71.
41. Ghatalia P, Gordetsky J, Kuo F, et al. Prognostic impact of immune gene expression signature and tumor infiltrating immune cells in localized clear cell renal cell carcinoma. *J Immunother Cancer*. 2019;7(1):139.
42. Wang Q, Zhou D, Wu F, et al. Immune Microenvironment Signatures as Biomarkers to Predict Early Recurrence of Stage Ia-b Lung Cancer. *Front Oncol*. 2021;11:680287.

## FIGURE LEGENDS

**Figure 1**. **Overview of tissue collection and analyses**. DCIS biopsies were processed for biomarker evaluation via gene expression microarrays, DNA sequencing, and staining tissue sections. The number of samples analyzed under each platform are shown in red.

**Figure 2. DCIS genomic landscape.** (**A**). Smoothed frequency (y-axis) of CNA gains (top) and losses (bottom) smoothed along the genome (x-axis) for DCIS. (**B**). Boxplot comparing percent of genome altered by copy number in DCIS cases with none, DCIS, or invasive recurrences (Wilcoxon test p values shown). (**C**) Dot matrix plot showing associations of select immune biomarkers with chromosomal gains (red) or losses (blue). Size of dot is proportional to significance (larger dot → smaller p-value). Dots are only shown for significant associations ($p<0.05$; Kruskal-Wallis test). (**D**) Oncoprint diagram displaying the mutational status of driver genes commonly altered in breast cancer. Clinical characteristics and percent TILs are indicated in the bottom annotations. (**E, F**) Boxplots showing associations of TILs with TP53 and PIK3CA mutations .

**Figure 3. Associations of biomarkers with clinical characteristics and outcomes of high-risk DCIS.**
Association dot-matrix showing the level and direction of association between clinical parameters and (**A**) immune or stromal biomarkers identified by multiplex IF or single-plex IHC and (**B**) gene expression signatures. Only those biomarkers that were significant ($p<0.05$) with at least one clinical variable are shown. Color of dot indicates direction of association (red – positive association; blue – negative association). Size of dot is proportional to significance (larger dot → smaller p-value); Kruskal-Wallis tests used for Grade comparisons; Wilcoxon tests used for Comedo, ClinPalp, PathMass, HR, and HER2 comparisons. **C-F**. Kaplan-Meier curves showing $CD3^{+}CD8^{-}$ T cells were associated with good outcomes for all recurrences (**C**), a Treg gene signature was associated with poor outcomes for all recurrences (**D**), and a proliferation gene signature was associated with shorter recurrence free interval for all recurrences (**E**) and for invasive recurrences (**F**). For the dichotomization of the biomarkers, median values were used as cut points (red=high; blue=low). LogRank p values are shown.

**Figure 4. Associations of spatial biomarkers with clinical characteristics and outcomes of high-risk DCIS.**
**(A)** Association dot-matrix showing the level and direction of association between clinical parameters and spatial biomarkers identified by multiplex IF. Only those biomarkers that were significant ($p<0.05$) with at least one clinical variable are shown. Color of dot indicates direction of association (red – positive association; blue – negative association). Size of dot is proportional to significance (larger dot → smaller p-value); Kruskal-Wallis tests used for Grade comparisons; Wilcoxon tests used for Comedo, ClinPalp, PathMass, HR, and HER2 comparisons. (**B,C)** Kaplan-Meier curves showing close spatial proximity of tumor cells and T cells (high Tum_T.SPS; red line) was associated with good outcomes for all recurrences (**B**) as well as invasive recurrences (**C**). (**D,E)** Kaplan-Meier curves showing low spatial proximity of tumor cells and T cells combined with a high proliferation score (blue line) was significantly associated with poor outcomes for all recurrences **(D)** as well as invasive recurrences **(E)**. For the dichotomization of the biomarkers, median values were used as cut points. LogRank p values are shown.

**Figure 5. Distribution and spatial proximity of aSMA/TP63 co-expressing cells within the myoepithelium.** (**A-F**) Representative micrographs showing distribution of aSMA/TP63 co-

expressing cells within the myoepithelium of different types of lesions: (**A-C**) benign; (**D-F**) DCIS; (**A,D**) all markers (DAPI, CD3, TP63, CD90, CD34, aSMA, CD68); (**B, E**) TP63/aSMA co-expressing cells within the myepithelium; (**C,F**) Markup image after analysis using HALO (Indica Labs) showing only TP63/aSMA co-expressing cells. Red arrowheads showing distribution of TP63/aSMA co-positive cells between benign and DCIS lesions. (**G**) Boxplot comparing myoepithelial continuity in DCIS regions compared to benign (ADH) and normal regions (Wilcoxon test p values shown). (**H**) Kaplan-Meier curves showing low myoepithelial continuity (blue line) was associated with better outcomes for all recurrences. **(I)** Kaplan-Meier curves showing low myoepithelial continuity combined with high percent of $CD3^{+}CD8^{-}$ T cells (blue line) was associated with good outcomes for all recurrences. For the dichotomization of the biomarkers, median values were used as cut points. LogRank p values are shown.

FIGURES

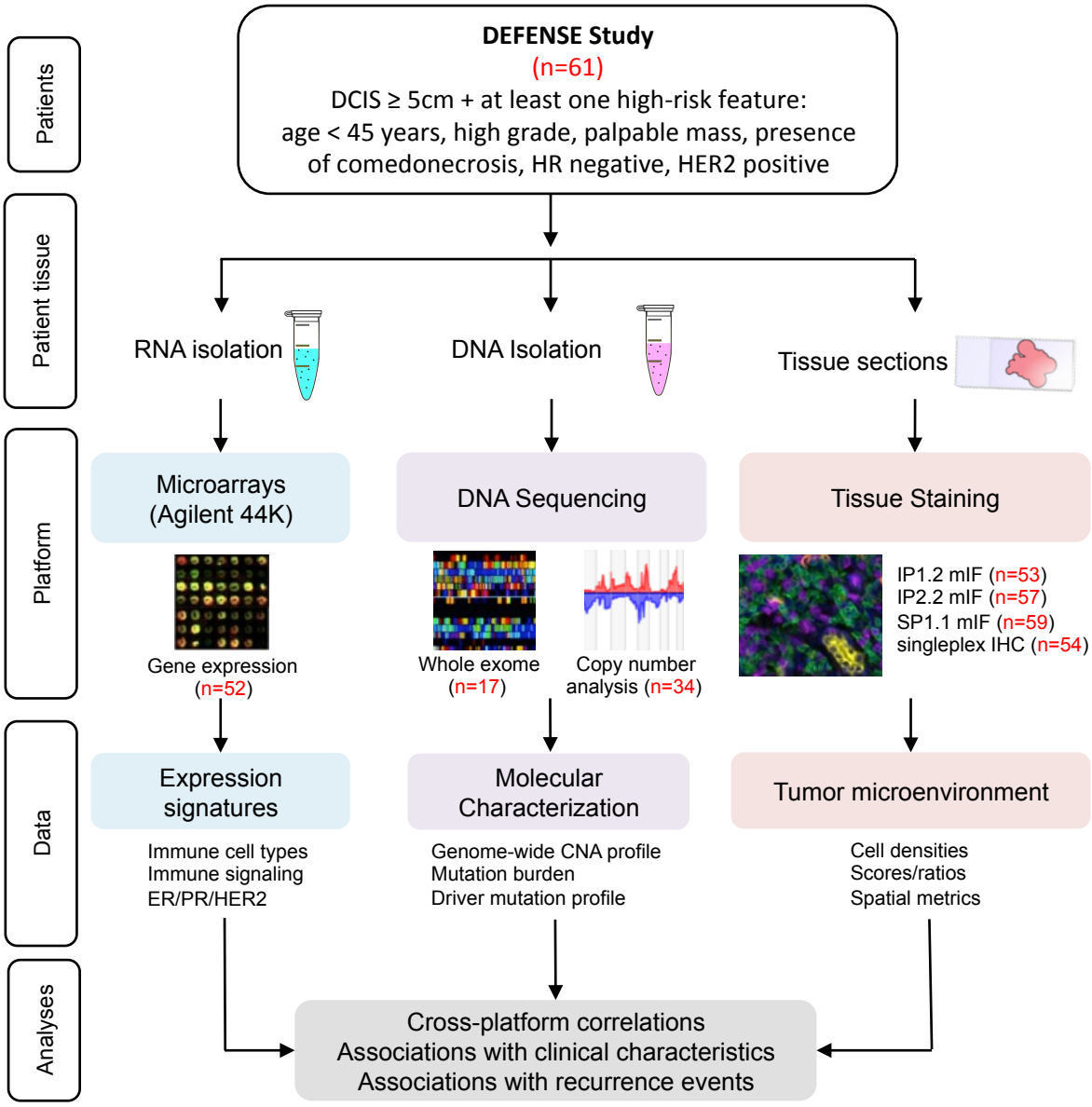


**Figure 1.**

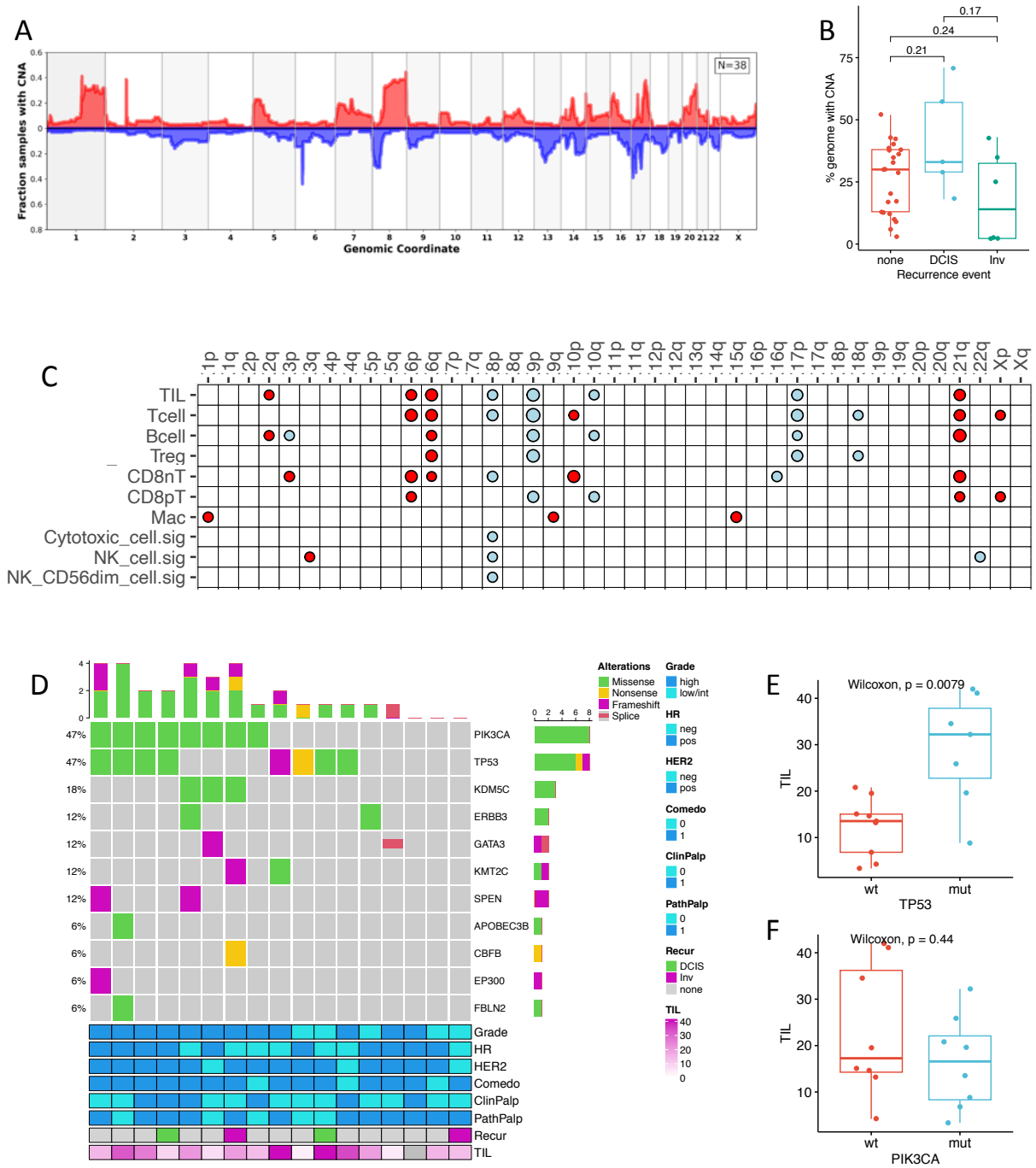
A
Fraction samples with CNA
Genomic Coordinate
N=38
B
% genome with CNA
0.17
0.24
0.21
none
DCIS
Inv
Recurrence event
C
TIL
Tcell
Bcell
Treg
CD8nT
CD8pT
Mac
Cytotoxic_cell.sig
NK_cell.sig
NK_CD56dim_cell.sig
D
PIK3CA
TP53
KDM5C
ERBB3
GATA3
KMT2C
SPEN
APOBEC3B
CBFB
EP300
FBLN2
Grade
HR
HER2
Comedo
ClinPalp
PathPalp
Recur
TIL
Alterations
Missense
Nonsense
Frameshift
Splice
E
Wilcoxon, p = 0.0079
TIL
wt
mut
TP53
F
Wilcoxon, p = 0.44
PIK3CA

**Figure 2.**

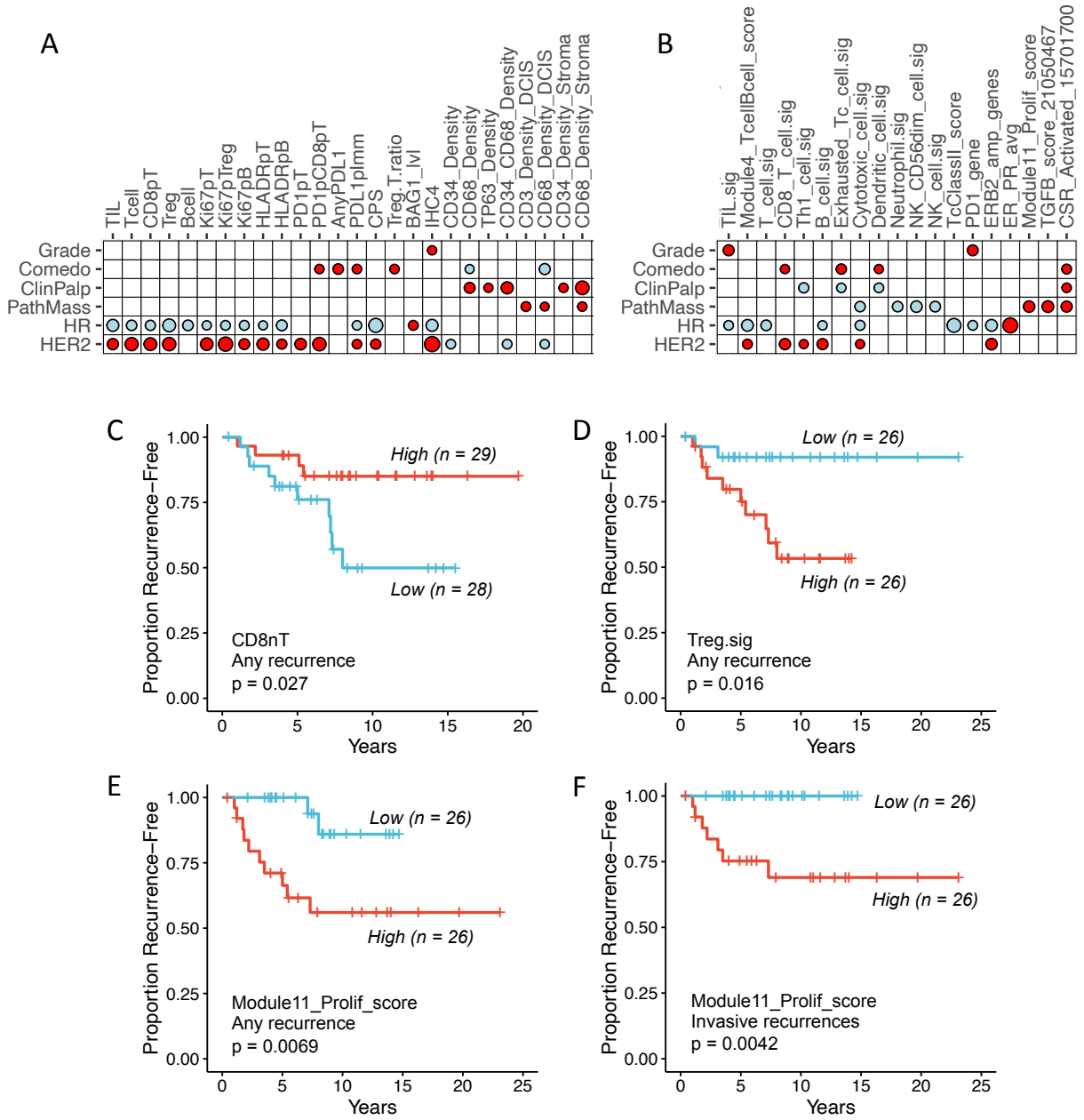
A
TIL
Tcell
CD8pT
Treg
Bcell
Ki67pT
Ki67pTreg
Ki67pB
HLADRpT
HLADRpB
PD1pT
PD1pCD8pT
AnyPDL1
PDL1plmm
CPS
Treg.T.ratio
BAG1_lvl
IHC4
CD34_Density
CD68_Density
TP63_Density
CD34_CD68_Density
CD3_Density_DCIS
CD68_Density_DCIS
CD34_Density_Stroma
CD68_Density_Stroma
Grade
Comedo
ClinPalp
PathMass
HR
HER2
B
TIL.sig
Module4_TcellBcell_score
T_cell.sig
CD8_T_cell.sig
Th1_cell.sig
B_cell.sig
Exhausted_Tc_cell.sig
Cytotoxic_cell.sig
Dendritic_cell.sig
Neutrophil.sig
NK_CD56dim_cell.sig
NK_cell.sig
TcClassII_score
PD1_gene
ERB2_amp_genes
ER_PR_avg
Module11_Prolif_score
TGFB_score_21050467
CSR_Activated_15701700
Grade
Comedo
ClinPalp
PathMass
HR
HER2
C
Proportion Recurrence−Free
High (n = 29)
Low (n = 28)
CD8nT
Any recurrence
p = 0.027
Years
D
Low (n = 26)
High (n = 26)
Treg.sig
Any recurrence
p = 0.016
E
Low (n = 26)
High (n = 26)
Module11_Prolif_score
Any recurrence
p = 0.0069
F
Low (n = 26)
High (n = 26)
Module11_Prolif_score
Invasive recurrences
p = 0.0042

**Figure 3.**

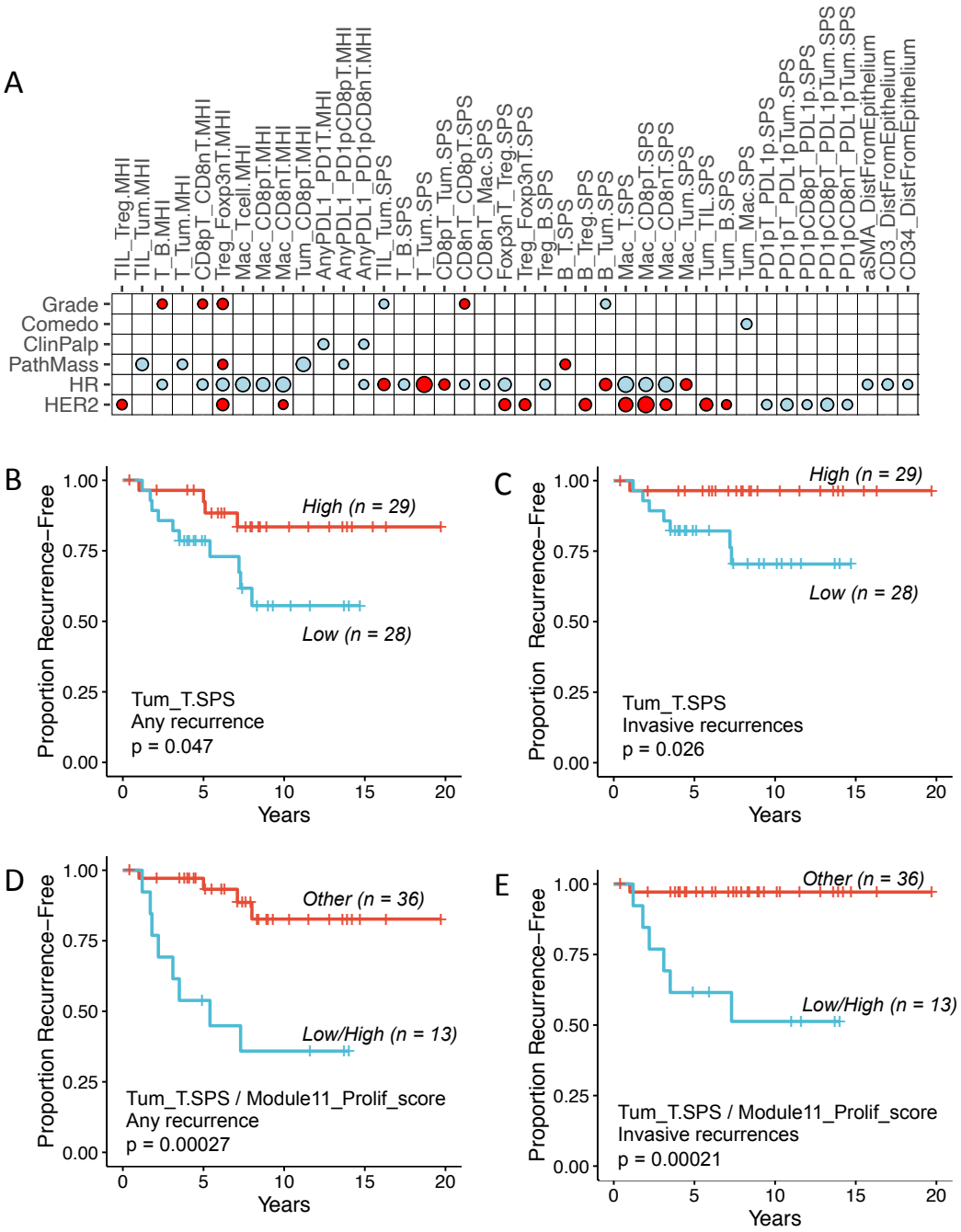
A
TIL_Treg.MHI
TIL_Tum.MHI
T_B.MHI
T_Tum.MHI
CD8pT_CD8nT.MHI
Treg_Foxp3nT.MHI
Mac_Tcell.MHI
Mac_CD8pT.MHI
Mac_CD8nT.MHI
Tum_CD8pT.MHI
AnyPDL1_PD1T.MHI
AnyPDL1_PD1pCD8pT.MHI
AnyPDL1_PD1pCD8nT.MHI
TIL_Tum.SPS
T_B.SPS
T_Tum.SPS
CD8pT_Tum.SPS
CD8pT_CD8pT.SPS
CD8nT_Mac.SPS
Foxp3nT_Treg.SPS
Treg_Foxp3nT.SPS
Treg_B.SPS
B_T.SPS
B_Treg.SPS
B_Tum.SPS
Mac_T.SPS
Mac_CD8pT.SPS
Mac_CD8nT.SPS
Mac_Tum.SPS
Tum_TIL.SPS
Tum_B.SPS
Tum_Mac.SPS
PD1pT_PDL1p.SPS
PD1pT_PDL1pTum.SPS
PD1pCD8pT_PDL1p.SPS
PD1pCD8pT_PDL1pTum.SPS
PD1pCD8nT_PDL1pTum.SPS
aSMA_DistFromEpithelium
CD3_DistFromEpithelium
CD34_DistFromEpithelium
Grade
Comedo
ClinPalp
PathMass
HR
HER2
B
Proportion Recurrence−Free
High (n = 29)
Low (n = 28)
Tum_T.SPS
Any recurrence
p = 0.047
Years
C
High (n = 29)
Low (n = 28)
Tum_T.SPS
Invasive recurrences
p = 0.026
D
Other (n = 36)
Low/High (n = 13)
Tum_T.SPS / Module11_Prolif_score
Any recurrence
p = 0.00027
E
Other (n = 36)
Low/High (n = 13)
Tum_T.SPS / Module11_Prolif_score
Invasive recurrences
p = 0.00021

**Figure 4.**

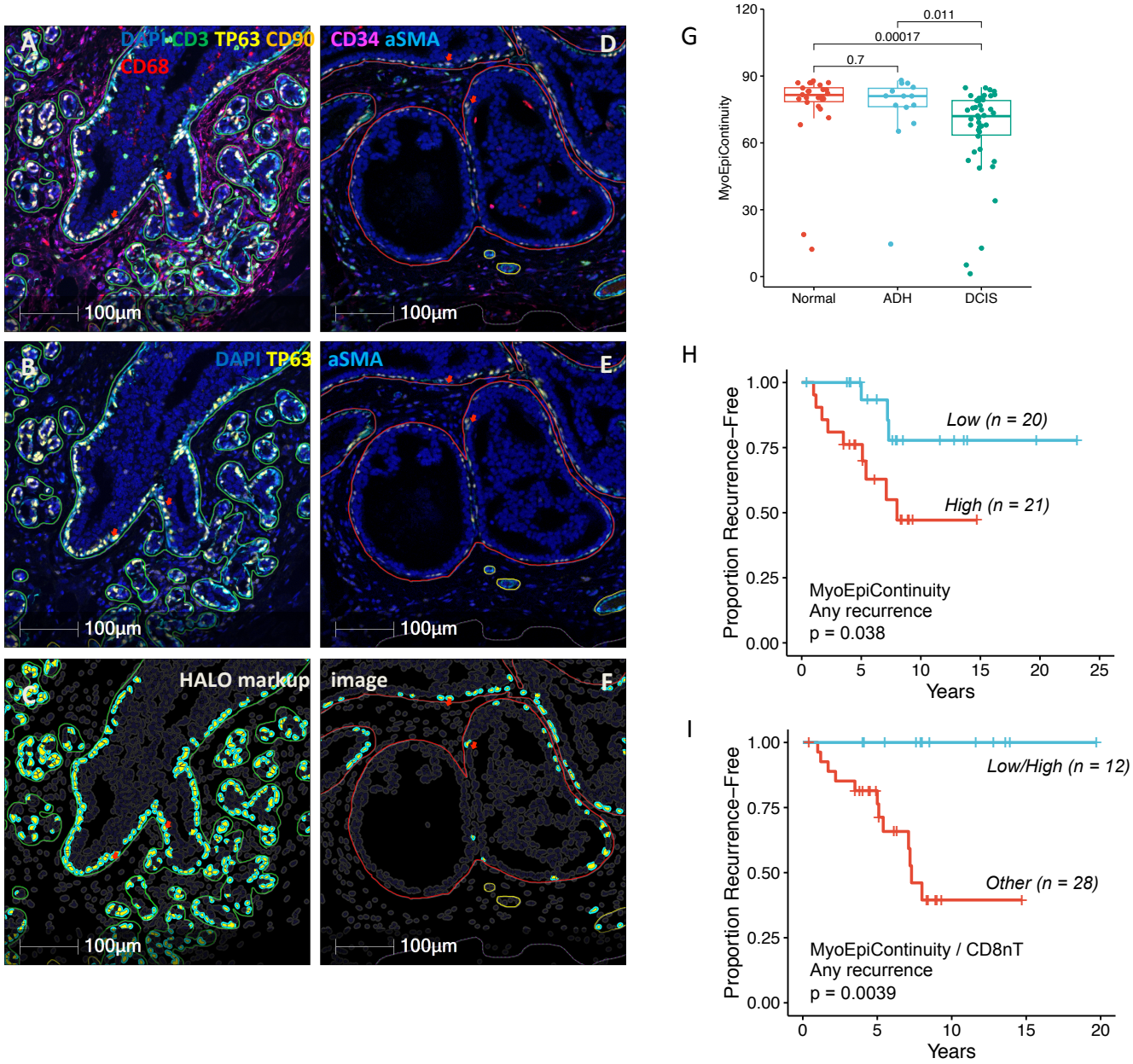
A
DAPI CD3 TP63 CD90
CD68
100µm
D
CD34 aSMA
100µm
B
DAPI TP63
100µm
E
aSMA
100µm
C
HALO markup
100µm
F
image
100µm
G
MyoEpiContinuity
0.011
0.00017
0.7
Normal
ADH
DCIS
H
Proportion Recurrence–Free
Low (n = 20)
High (n = 21)
MyoEpiContinuity
Any recurrence
p = 0.038
Years
I
Proportion Recurrence–Free
Low/High (n = 12)
Other (n = 28)
MyoEpiContinuity / CD8nT
Any recurrence
p = 0.0039
Years

**Figure 5.**