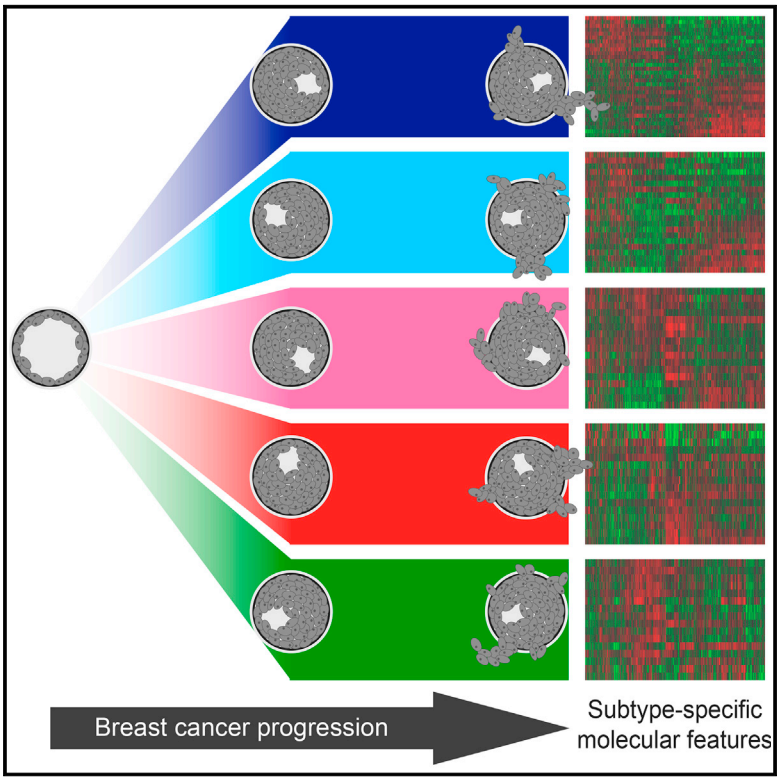
Cell Reports

Resource

Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer

Graphical Abstract



Authors

Robert Lesurf, Miriam Ragle Aure, Hanne Håberg Mørk, ..., Fredrik Wärnberg, Michael Hallett, Therese Sørlie

Correspondence

therese.sorlie@rr-research.no

In Brief

Lesurf et al. show that progression from DCIS to invasive breast cancer occurs in an intrinsic subtype-specific manner, indicating distinct evolutionary disease paths. Most published signatures are consistently biased for improved performance in the luminal subtypes. Subtype-specific molecular markers may have the potential to predict risk of progression.

Highlights

- Gene expression differences between DCIS and IDC are associated with intrinsic subtype
- Each subtype likely undergoes a distinct evolutionary course of disease progression
- Molecular properties that differentiate DCIS and IDC involve the microenvironment
- Subtype-specific signatures may have the potential to predict invasiveness

Accession Numbers

GSE59248

Lesurf et al., 2016, Cell Reports *16*, 1166–1179
July 26, 2016
http://dx.doi.org/10.1016/j.celrep.2016.06.051

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Molecular Features of Subtype-Specific Progression from Ductal Carcinoma In Situ to Invasive Breast Cancer

Robert Lesurf,[1,2,3] Miriam Ragle Aure,[4,5] Hanne Håberg Mørk,[4] Valeria Vitelli,[6] Oslo Breast Cancer Research Consortium (OSBREAC), Steinar Lundgren,[7,8] Anne-Lise Børresen-Dale,[4,5] Vessela Kristensen,[4,5,9] Fredrik Wärnberg,[10] Michael Hallett,[1,2,11] and Therese Sørlie[4,5,*]

[1]Department of Biochemistry, McGill University, Montreal, QC H3G 1Y6, Canada
[2]McGill Centre for Bioinformatics, McGill University, Montreal, QC H3G 1Y6, Canada
[3]Rosalind & Morris Goodman Cancer Research Centre, McGill University, Montreal, QC H3A 1A3, Canada
[4]Institute for Cancer Research and Department of Cancer Genetics, Oslo University Hospital, The Norwegian Radium Hospital, 0424 Oslo, Norway
[5]K.G. Jebsen Centre for Breast Cancer Research, Institute for Clinical Medicine, University of Oslo, 0318 Oslo, Norway
[6]Oslo Center for Biostatistics and Epidemiology and Department of Biostatistics, University of Oslo, 0317 Oslo, Norway
[7]Department of Oncology, St. Olav's University Hospital, 7006 Trondheim, Norway
[8]Department of Cancer Research and Molecular Medicine, Faculty of Medicine, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway
[9]Division of Medicine, Department of Clinical Molecular Biology and Laboratory Science (EpiGen), Akershus University Hospital, 1478 Lørenskog, Norway
[10]Department of Surgical Sciences, Uppsala University, 751 85 Uppsala, Sweden
[11]School of Computer Science, McGill University, Montreal, QC H3A 0E9, Canada
*Correspondence: therese.sorlie@rr-research.no
http://dx.doi.org/10.1016/j.celrep.2016.06.051

SUMMARY

Breast cancer consists of at least five main molecular "intrinsic" subtypes that are reflected in both pre-invasive and invasive disease. Although previous studies have suggested that many of the molecular features of invasive breast cancer are established early, it is unclear what mechanisms drive progression and whether the mechanisms of progression are dependent or independent of subtype. We have generated mRNA, miRNA, and DNA copy-number profiles from a total of 59 in situ lesions and 85 invasive tumors in order to comprehensively identify those genes, signaling pathways, processes, and cell types that are involved in breast cancer progression. Our work provides evidence that there are molecular features associated with disease progression that are unique to the intrinsic subtypes. We additionally establish subtype-specific signatures that are able to identify a small proportion of pre-invasive tumors with expression profiles that resemble invasive carcinoma, indicating a higher likelihood of future disease progression.

INTRODUCTION

Approximately 20% of all breast cancer detected through mammography are ductal carcinoma in situ (DCIS), a pre-invasive form of the disease (Ernster et al., 2002). However, it has been estimated that 20%–50% of DCIS tumors would progress to invasive ductal carcinoma (IDC) if left untreated (Sanders et al., 2005). Nonetheless, the natural history of DCIS and the exact causes of disease progression are unknown, and we lack accurate ways to identify at diagnosis those DCIS patients that may be safely spared treatment (Cowell et al., 2013; Espina et al., 2010; Falk et al., 2011; Kaur et al., 2013). A recent observational study of more than 100,000 women found that the breast cancer specific mortality rate following a DCIS diagnosis was about 3%, emphasizing a need for markers that predict disease progression (Narod et al., 2015).

Both in situ and invasive breast tumors are comprised of heterogeneous phenotypes, with variation in clinicopathological features such as histological grade, estrogen receptor alpha (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2/ERBB2), and epidermal growth factor receptor (EGFR) status (Clark et al., 2011; Livasy et al., 2007). Additionally, tumors of the different intrinsic subtypes are found among both DCIS and IDC (Allred et al., 2008; Hannemann et al., 2006; Muggerud et al., 2010; Vincent-Salomon et al., 2008), indicating subtype-specific progression paths. DCIS lesions are commonly found to have copy-number alterations that are characteristic of the invasive disease, suggesting that genomic instability is an early event in breast tumorigenesis (Berman et al., 2005; Chin et al., 2004). Moreover, previous studies have also suggested that clinical and intrinsic subtype, as opposed to disease stage, are the dominant sources of variability among tumor expression profiles (Hannemann et al., 2006; Muggerud et al., 2010).

Nonetheless, differences do exist in the proportions of lesions within each disease stage that carry specific genomic features. For example, a higher proportion of DCIS have amplification of *ERBB2* (Park et al., 2006; van de Vijver et al., 1988), while *FGFR1* (Jang et al., 2012) and *MYC* (Robanus-Maandag et al., 2003) amplifications are more frequent in the invasive stage. Among genomic studies, analyses in both bulk and microdissected tissues have suggested that those genes differentially expressed between DCIS and IDC are functionally enriched for changes in cell types and processes relating to the tumor microenvironment (Lee et al., 2012; Ma et al., 2003, 2009; Vargas et al., 2012). Similarly, others have provided evidence that epigenetic changes contribute to breast cancer progression (Fazzari and Greally, 2004; Fleischer et al., 2014; Widschwendter and Jones, 2002) and that the process may be mediated by myoepithelial cells (Hu et al., 2008).

We previously identified a distinct subgroup of DCIS tumors with characteristics of invasive disease and from this developed a signature for the detection of similar tumors in independent data (Muggerud et al., 2010). Consistent with other studies, we found that high grade DCIS tumors are more likely than low grade tumors to progress to the invasive disease (Fitzgibbons et al., 1998; Hughes et al., 2009; Sanders et al., 2005) and suggest that there are molecular properties present in some DCIS that may be predictive of disease progression. Since tissues tend to co-cluster by intrinsic subtype and not disease stage (e.g., basal-like tumors are more highly correlated with other basal-like tumors regardless of whether they are IDC or DCIS), studies that investigate the molecular determinants of progression must incorporate subtype in a careful manner.

To address these issues, we have investigated the molecular profiles of DCIS, lesions with a mixed diagnosis, small IDC tumors using gene and microRNA (miRNA) expression analysis, and SNP arrays for estimating DNA copy number. We identify molecular features of progression that are unique to each subtype. Moreover, we demonstrate that features that differentiate disease stage in an unstratified analysis are systematically confounded by subtype. Our analysis further identifies a small number of DCIS lesions with profiles suggesting they may be more likely to progress to an invasive state. The features that differentiate these DCIS from indolent DCIS may allow for the identification of similar lesions at the time of diagnosis, which may in turn have implications for treatment decisions.

RESULTS

DCIS and IDC Molecular Profiles Cluster According to Intrinsic Subtype, but Signatures of Disease Stage Are Observed within Subtype Clusters

To investigate the heterogeneity of mRNA profiles of DCIS and IDC, we filtered for the most variable genes and performed class discovery (see the Supplemental Information; Figure S1A). Samples divided into two main clusters primarily by ER status and luminal characteristics. The intrinsic subtypes were differentially enriched between the two clusters (Table S1, Fisher's exact test, and p value $< 2 \times 10^{-10}$). This was also true for ER-positive samples (p value $< 4 \times 10^{-11}$), HER2-positive samples (p value $< 3 \times 10^{-2}$), and samples of distinct grades (p value $< 5 \times 10^{-2}$). However, neither cluster was enriched for the DCIS or IDC disease stage, nor for lymph node-positive cases (p value > 0.05). Within each of the two primary clusters, subclusters consisting of only DCIS or IDC were observed. Together, this indicates that although disease stage is not the primary source of variability across breast gene expression data, systematic differences are observed between DCIS and IDC within the main subgroups of lesions.

Interestingly, analogous unsupervised analysis using the miRNA data did not identify similar associations between primary clusters and clinicopathological features (see the Supplemental Information; Figure S1B). For the miRNA data, the two primary clusters (i and ii) were enriched for significant differences in disease stage (Table S1, p value $< 2 \times 10^{-2}$), but there were no significant associations with intrinsic subtype, ER status, HER2 status, lymph node status, or grade (p value > 0.05). However, when comparing primary cluster ii with subclusters iii and iv, significant differential enrichments were found with respect to the type of lesion (p value $< 6 \times 10^{-3}$) and intrinsic subtype (p value $< 4 \times 10^{-2}$). Correspondingly, we noted that cluster iii contains only a single non-luminal lesion and that cluster iv contains only a single IDC. No associations were found for HER2 status, lymph node status, or grade.

Previously reported copy-number alteration events in breast cancer were observed in multiple samples regardless of disease stage. These include gains in chromosome 1q and 8q and losses in chromosome 8p and 16q. As expected, high grade samples had a higher frequency of copy-number alterations than low grade samples. We confirmed this trend by calculating the genomic grade index (GGI) (Chin et al., 2007) of the samples. Grade III tumors were found to have significantly higher GGI scores than the combined grade I and II scores (Figure S2A; p value < 0.02). Similarly, samples classified as NormL had the lowest GGI scores, and those classified as BasalL had the highest GGI scores (Figure S2B; p value $< 9 \times 10^{-4}$).

Tumors primarily clustered according to their ploidy (Figure S3, clusters i and ii), with all samples in cluster i having apparent whole genome duplications. We did not observe significant enrichments in these two primary clusters for stage of lesion or other clinicopathological features (intrinsic subtype, ER, HER2, lymph node, and grade; Table S1, p value > 0.05). However, when comparing cluster i with subclusters iii and iv, significant differences for enrichment were found between intrinsic subtype (p value $< 5 \times 10^{-2}$) and grade (p value $< 4 \times 10^{-3}$). No other significant associations were found for clinicopathological features.

In summary, at each data level, sample clusters enriched for intrinsic subtype were observed. This was a particularly dominant signal at the gene expression level, where the subtypes were originally defined. In order to determine whether combining different data levels results in the identification of sample groups systematically associated with different disease stages, we calculated the IntClust subtypes (Curtis et al., 2012). Samples belonging to each of the IntClust subtypes were observed, with the exception of IC2 (Table S1). Overall, no significant differences were identified between the proportion of DCIS and IDC assigned to each IntClust subtype (multidimensional Fisher's exact test and p value > 0.05). Similar findings were observed

when investigating the proportion of DCIS and IDC assigned to each transcriptionally derived intrinsic subtype (Table S1).

Intrinsic Subtypes Have Confounded Previous Attempts to Differentiate Pre-invasive and Invasive Disease Stages

Although breast lesions do not primarily divide according to the disease stage, we were nonetheless able to identify 188 genes differentially expressed between DCIS versus IDC (Figure 1A; limma, adjusted p value < 0.05). This set was enriched for processes such as collagen fibril organization, extracellular matrix interactions, and focal adhesion, which are characteristic of cells in the microenvironment (Table S1). This is in line with the types of processes that have previously been identified as discriminating bulk expression profiles of DCIS from IDC. Furthermore, our set of differentially expressed genes generally had a significant degree of overlap with similarly derived signatures that have previously been published (Hannemann et al., 2006; Knudsen et al., 2012; Lee et al., 2012; Ma et al., 2003, 2009; Porter et al., 2003; Schuetz et al., 2006) (one-sided Fisher's exact test, Table S1).

Most published DCIS versus IDC signatures, and our list of 188 genes, contain at least one gene from the PAM50 gene set used for intrinsic subtyping. Often the degree of overlap was statistically significant (one-sided Fisher's exact test, Table S1).

The surprisingly high overlap between purported signatures of DCIS versus IDC and classification markers of subtype motivated a more detailed examination. Toward this end, we derived a gene signature for each of the breast cancer intrinsic subtypes (see the Supplemental Information). We compared the set of genes that classified each subtype to our DCIS versus IDC gene list and found that every subtype contributed a surprising number of genes (Fisher's exact test, LumA p value $< 3 \times 10^{-12}$, LumB p value $< 2 \times 10^{-4}$, Her2E p value $< 9 \times 10^{-15}$, BasalL p value $< 3 \times 10^{-14}$, and NormL p value $< 5 \times 10^{-17}$).

One may expect that the genes that are differentially expressed between DCIS and IDC would be part of molecular processes or cellularity that are ubiquitous across subtypes. In fact, our signature demonstrated significantly different levels of activation between the subtypes (Figure 1B). Additionally, our DCIS versus IDC gene list failed to separate DCIS from IDC within the NormL subtype (Figure 1C). Together, this suggests that frequently cited genes and signatures related to progression from a pre-invasive to an invasive state are systematically incorrect for certain breast cancer subtypes.

In contrast to the mRNA data, we did not identify any miRNA as differentially expressed between DCIS and IDC (limma, adjusted p value < 0.05). The Volinia (Volinia et al., 2012) signature of nine miRNAs that were found to be differentially expressed between eight DCIS and 80 IDC tumors did not separate DCIS lesions from IDC in our study, although four of these miRNAs (miR-126, miR-143, miR-218, and miR-221) were differentially expressed by raw p value (see Supplemental Information; Figure S4).

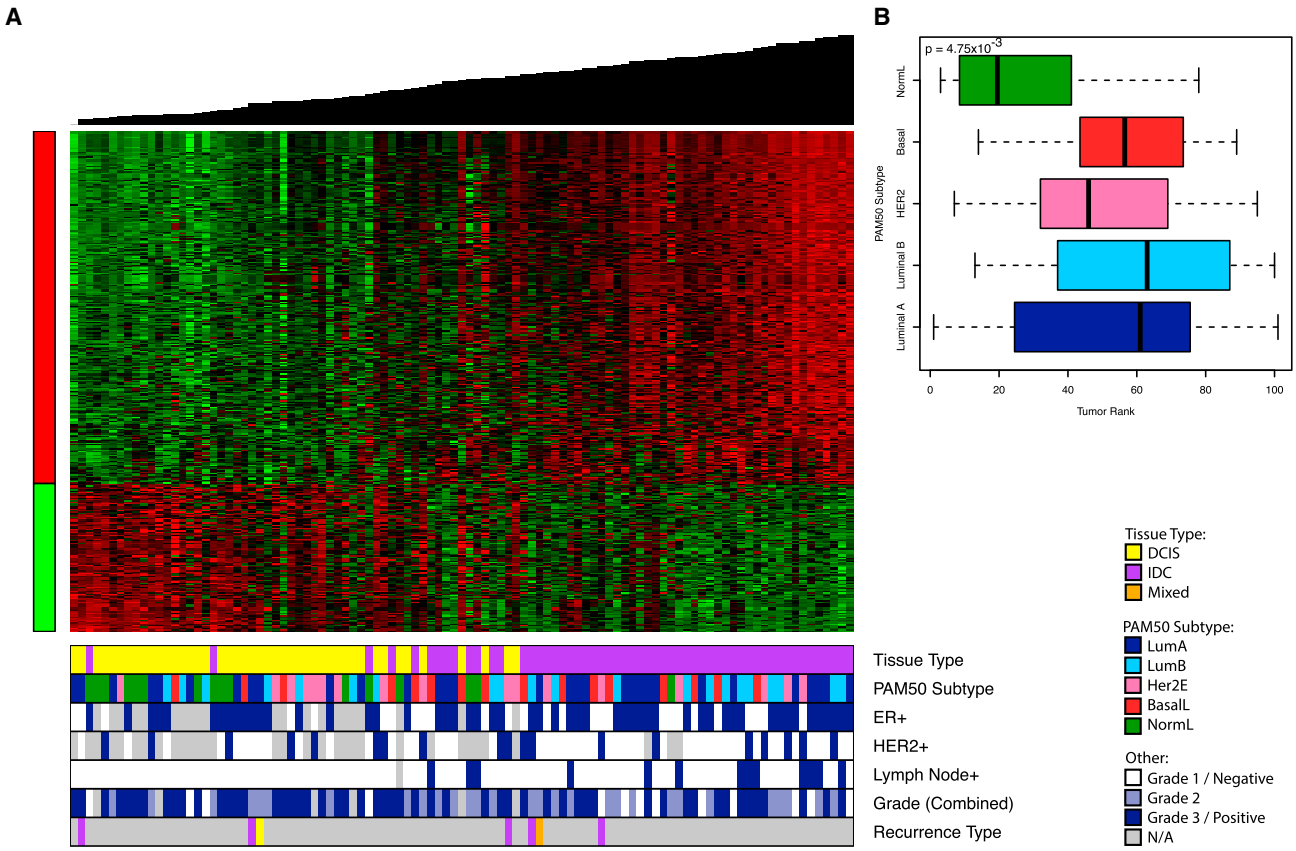
We did not identify any genomic loci for which there were consistent differential copy-number changes between DCIS and IDC. Overall, relative copy-number gains of *ERBB2* were proportionately similar in all disease stages (3/39 DCIS, 4/64 IDC, and 2/16 mixed type; Fisher's exact test and p value > 0.05). However, there was a trend toward a relative gain in *FGFR1* copy-number in invasive tumors (6/39 DCIS, 20/64 IDC, and 2/16 mixed type; p value $< 2 \times 10^{-2}$). Similarly, there was a relative gain in *MYC* copy number in invasive tumors (9/39 DCIS, 30/64 IDC, and 5/16 mixed type; p value $< 5 \times 10^{-2}$), consistent with previous studies (Jang et al., 2012; Robanus-Maandag et al., 2003). Interestingly, *MYC* copy-number gains were more significantly associated with the PAM50 subtype than with disease stage (p value $< 8 \times 10^{-4}$).

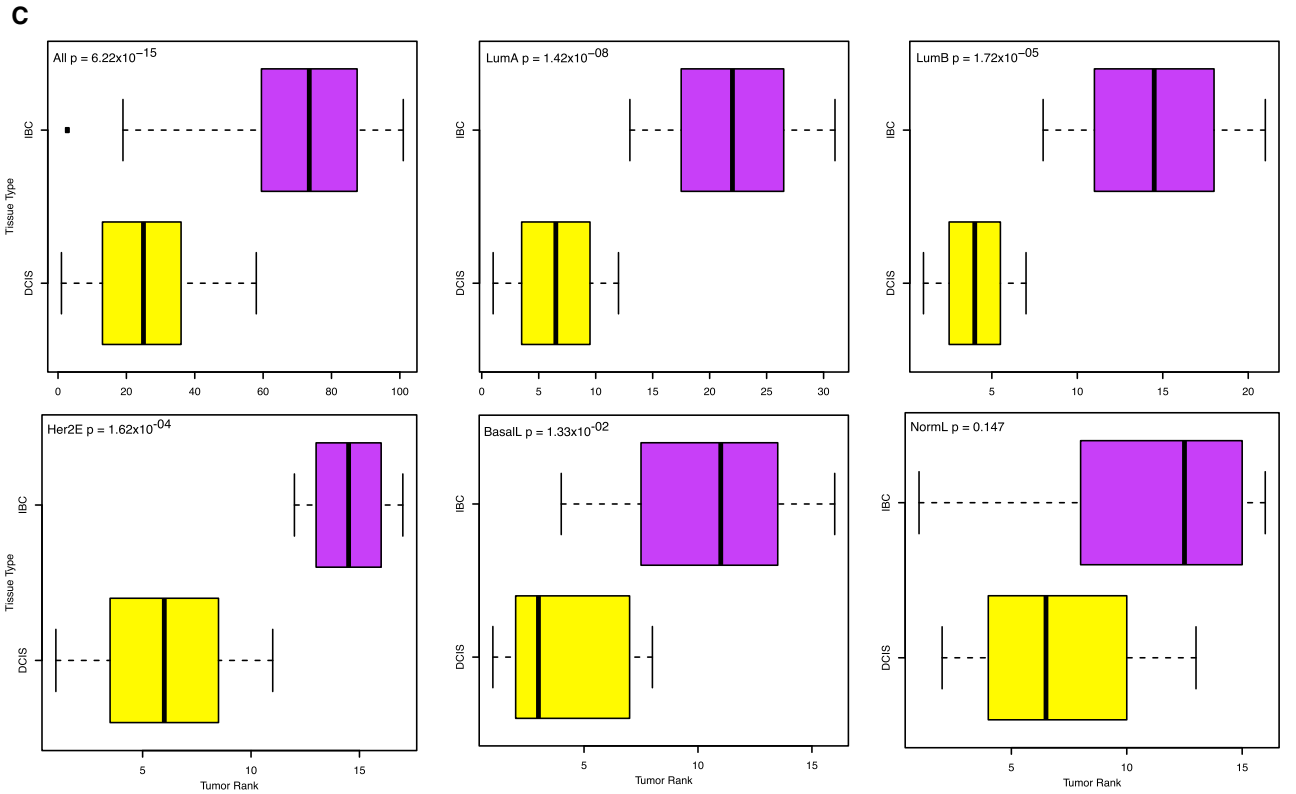
We additionally collected sequencing data on *TP53* and *PIK3CA*, the two genes most frequently mutated in breast cancer (Cancer Genome Atlas Network, 2012). Here, we sought to determine whether these mutation rates differ between DCIS and IDC, possibly as a result of the accumulation of de novo mutations during disease progression. However, the mutation rates were similar between DCIS and IDC for both genes, supporting the notion that such mutations are acquired at an earlier disease stage than DCIS (Table S1, Fisher's exact test, and p value > 0.05).

Subtype-Specific Measures of Progression Are Disjoint

Given the confounding we observed between histopathological variables, intrinsic subtype, and disease stage, we generated a linear model that incorporates the variables of ER status, HER2 status, tumor grade, and intrinsic subtype into the identification of differentially expressed genes between DCIS versus IDC (see the Supplemental Information). The model identified six significantly differentially expressed genes: *AS3MT*, *FAM74A4*, *GPR155*, *PQLC2*, *SLC16A12*, and *ZNF865* (limma, adjusted p value < 0.05). Interestingly, none of these genes were identified in our previous univariate analysis (Table S1), further indicating that the identification of genes that differentiate disease stage is confounded by clinicopathological variables and subtype.

Although this multivariate modeling approach was able to identify a small number of genes differentially expressed independently of intrinsic subtype, variables such as ER status are known to affect the transcriptional levels of thousands of genes, with complicated interactions with other variables such as the proliferative index of the tumor. As such, a simple linear model may not suffice to ablate the effects of these ubiquitous confounders. Patient stratification by subtype provides an alternative approach that potentially handles ubiquitous effects more appropriately. Under the hypothesis that DCIS lesions of any given subtype are most likely to progress to invasive tumors of the same subtype, we repeated our supervised class distinction approach by contrasting only DCIS versus IDC of the same subtype. Because there are a much smaller number of samples in each of these analyses in comparison to the unstratified analysis, we anticipated having less power, which in turn affects the number of features meeting a given significance threshold (Stretch et al., 2013). In order to compensate for these differences, we used a false discovery rate (FDR) cutoff of less than 0.05 for the within subtype analyses. The resultant gene lists are highly distinct across the subtypes, sharing no genes in common between all of them (Figures 2A and 2B; Table S1). Additionally, these lists differed in their ability to split DCIS from IDC. This was especially apparent among the LumA, LumB, and Her2E





(legend on next page)

subtypes, slightly less consistent with the BasalL subtype, and entirely untrue for the NormL subtype, where DCIS and IDC samples were always intermixed. The differences in the size of these gene lists (775 LumA, 90 LumB, 227 Her2E, 120 BasalL, and 0 NormL) did not appear to be directly related to the number of samples in the associated comparisons (12 versus 19 LumA, seven versus 14 LumB, 11 versus six Her2E, five versus 11 BasalL, and ten versus six NormL). They also did not appear to be related to the degree of heterogeneity within each subtype, as witnessed by similarities in the interquartile range distributions across all genes on the array, within each subtype (Figure S5A).

To further investigate the relationship between disease stage classification signatures and their within-subtype performance, we additionally obtained signatures relating to disease progression in breast cancer from both literature and database sources. The majority of these signatures were generated by directly contrasting gene expression of DCIS versus IDC. Across the unstratified and subtype-stratified cohorts, we used naive Bayes classifiers with a leave-one-out cross-validation strategy to determine how well each signature is able to predict disease stage (Figure 3; Table S1, see the Supplemental Information). A small number of samples were consistently incorrectly predicted for all signatures; in particular, NormL samples tended to always be classified as DCIS. In general, the signatures with the best performance differed in each subtype, although our 188 gene signature that differentiates all DCIS from IDC within our unstratified cohort, labeled ''Lesurf 2016 (unstratified)'', was among the best performing signatures in all subtypes. Moreover, the proportion of signatures that significantly predicted disease stage (Fisher's exact test, p value < 0.05) varied by cohort stratification, with 97% of signatures significant in the unstratified cohort and only 8% significant within the NormL subtype.

Biological Processes that Distinguish DCIS from IDC Are Related to the Microenvironment

Next, we performed hierarchical clustering across the unstratified cohort using the union of differentially expressed genes from our unstratified and subtype-specific lists. We observed four main clusters of genes, with various degrees of overlap between subtypes (denoted i–iv; Figure 4). Gene Ontology (GO) enrichment analysis revealed that each cluster is defined by biologically distinct processes (Table S1). Cluster i was enriched for metabolic processes, ii for pathways relating to cell development and structure, iii for various forms of immune response, and iv for genes related to the extracellular matrix.

Several tools were used to additionally investigate the pathways and processes that differentiate DCIS from IDC within each intrinsic subtype (Table S1, see the Supplemental Information). Overall, we found that the LumA, LumB, and Her2E subtypes shared many common pathways, especially those related to collagen fibril organization, cell adhesion, and ECM-cell receptor interactions. They also tended to share pathways that were differential between DCIS and IDC in the unstratified analysis. Nevertheless, there were also pathways specific to each of these subtypes, including cell differentiation and inflammatory immune responses for LumA, cell migration signatures for LumB, and enrichment for cell-cycle signatures in Her2E. The BasalL subtype was primarily enriched for properties related to the immune response, particularly with T cells, but there was also some evidence of B cell, immunoglobulin A, and natural killer cell presence. Finally, within the NormL subtype, no genes or gene sets were found to be differentially expressed between DCIS and IDC.

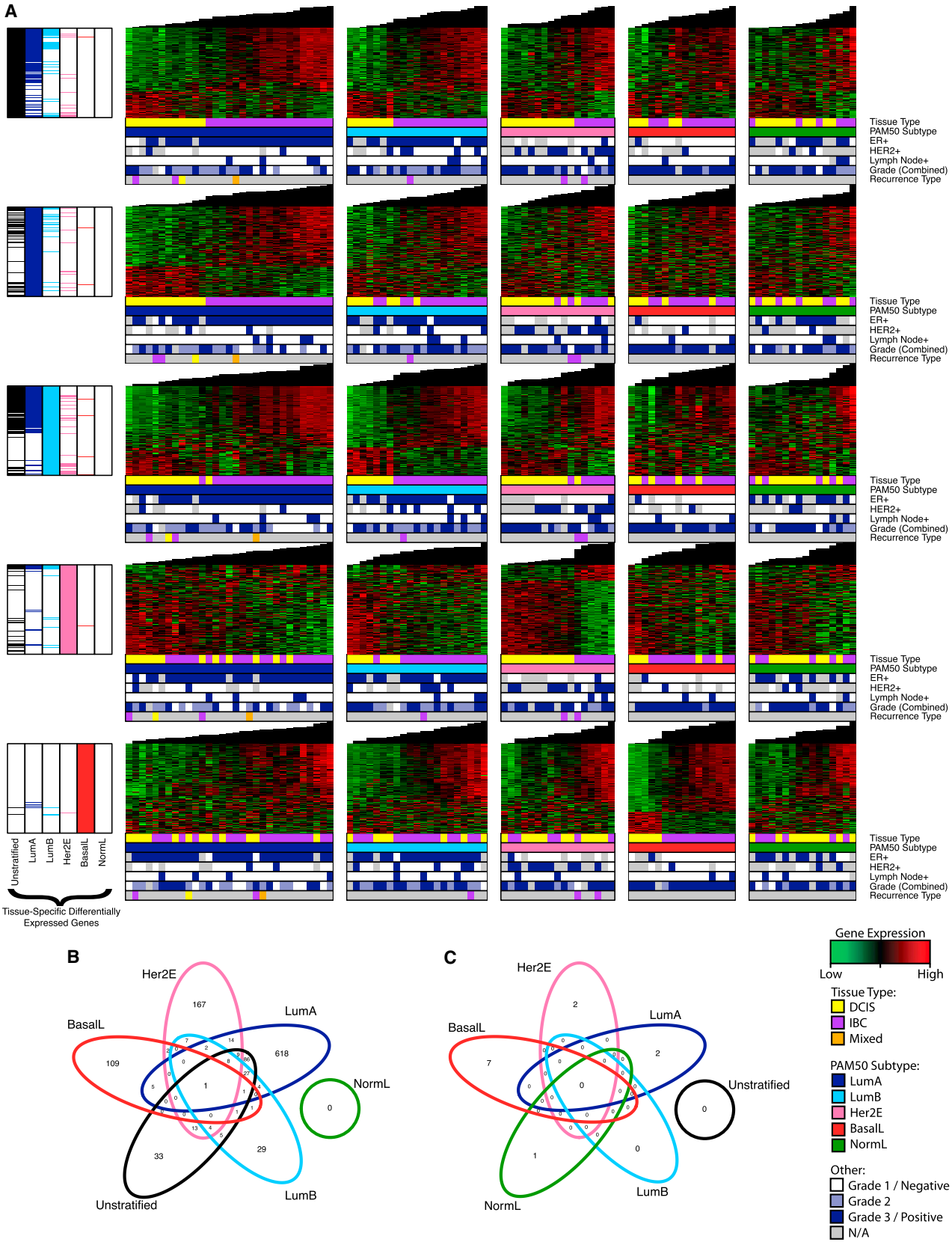
To further determine the role of the microenvironment in differentiating pre-invasive from invasive lesions, we investigated whether the genes differentially expressed between DCIS versus IDC also tend to be differentially expressed between epithelial and stromal tissue. Using two previously published data sets containing microdissected cell compartments from pre-invasive and invasive lesions (Lee et al., 2012; Ma et al., 2009), we found that in almost all cases there was significant overlap between our DCIS versus IDC gene lists and gene lists that differentiate the epithelial tissue from the stromal compartment of both DCIS and IDC (limma, FDR < 0.05 followed by a Fisher's exact test). The LumA subtype had the most significant overlaps (p value < 2×10^{-19} and 2×10^{-82} in the Lee and Ma lists, respectively) followed by Her2E (p values < 0.03 and < 6×10^{-05}). The BasalL and LumB subtypes were significantly only in the Ma list (p values < 0.02). Additionally, the list of genes differentially expressed between DCIS and IDC in the unstratified analysis was also significantly associated with the tumor microenvironment (p values < 5×10^{-4} and 1×10^{-25} for Lee and Ma lists, respectively).

Next, we asked whether there were miRNAs that were differentially expressed between DCIS and IDC per intrinsic subtype (Figure 2C). Only two miRNAs were identified for the LumA subtype (miR-10a* and miR-323-3p, both higher in IDC) and two for the Her2E subtype (miR-298 and miR-4300, both higher in IDC). Six miRNAs were higher in BasalL DCIS compared to IDC (miR-34c-5p, miR-95, miR-133b, miR-192, miR-218, and miR-363), one was lower (miR-K12-5*), and one miRNA (miR-136) was higher expressed in IDC compared to DCIS for the NormL subtype.

As with the earlier unstratified analysis, we were unable to identify systematic differences in copy-number or mutation rates of *TP53* and *PIK3CA* between the DCIS and IDC of each intrinsic subtype (Table S1). This is consistent with previous data demonstrating that changes at the DNA level occur early in breast tumorigenesis (Newburger et al., 2013).

Figure 1. Classification Signature of Breast Disease Stage

(A) Genes differentially expressed between DCIS and IDC. The upregulated and downregulated genes in IDC are denoted by red and green bars on the left side of the heatmap, respectively. The tumors are ordered according to an increasing rank-sum, and rank-sum scores are plotted as black bars on top of the heatmap.
(B) Boxplots representing tumor ranks stratified by each intrinsic subtype with respect to the signature shown in (A).
(C) Boxplots representing tumor ranks stratified by disease stage type. The ranks for all patients are shown in the upper left box and for each separate intrinsic subtype in the other boxes.



(legend on next page)

Predicting DCIS Patients that Are Most Likely to Progress

Although it is believed that not all DCIS lesions will progress to an invasive stage even in the absence of treatment, we still lack a means to distinguish between such indolent and aggressive DCIS. To this end, we have sought to identify DCIS lesions with invasive-like properties, under the hypothesis that these properties may be used as a classifier for likely progressors.

Under a leave-one-out cross-validation strategy within each subtype, we performed class distinction between IDC and DCIS after removing the single DCIS sample (limma, FDR < 0.05). Then using these genes as a basis set, we applied hierarchical clustering. The samples that clustered closest were compared to the true lesion type of the left-out patient. This process identified two LumA, one LumB, one Her2E, and one BasalL DCIS that clustered among the IDC; no NormL DCIS were identified, due to a lack of differentially expressed genes in each analysis (Figure 5A). These numbers are roughly concordant with estimates of indolent/aggressive DCIS from the literature (Page et al., 1982, 1995; Sanders et al., 2005). Although this investigation was not designed to study prognosis due to small sample size and selection of cases, we noted that one of these two ''invasive-like'' LumA DCIS lesions is known to have locally recurred, and similarly the only patient with an ''invasive-like'' Her2E DCIS has developed a distal invasive recurrence, suggesting the approach may in fact have efficacy to identify in situ tumors likely to progress.

For each of the intrinsic subtypes, we identified the genes that were significantly differentially expressed between those ''invasive-like'' DCIS and the remaining indolent samples and further filtered these lists by how consistently the genes were up- or downregulated in both IDC and ''invasive-like'' DCIS (Table S1, see the Supplemental Information). In total, 1,667 genes were identified across these four lists (1,283 LumA, 79 LumB, 261 Her2E, and 131 BasalL). We term these gene sets ''invasive signatures''. Although there was a small degree of overlap between the invasive signatures across the subtypes, there were not any genes common to all subtypes (Figure S5B). Interestingly, the invasive signatures were often enriched in pathways/processes that were distinct from the pathways/processes identified as differential between DCIS and IDC. The LumA invasive signature was primarily enriched for immune-related processes, as well as cell metabolism and cell cycle. The LumB invasive signature was also enriched for immune-related processes, including inflammatory responses, as well as cell-cell signaling. The Her2E invasive signature was enriched for cell adhesion, ECM-receptor interaction, cell motility, and cell morphogenesis. Finally, the BasalL invasive signature was primarily enriched for immune-related processes, predominantly those involving cytokine and inflammatory responses.

The invasive signatures were applied to previously published gene expression data sets in order to determine whether we could identify invasive-like DCIS. Notably, however, these data sets were smaller in size than our own, making it difficult to discern invasive patterns. Nonetheless, we were able to identify a small number of LumB and BasalL DCIS samples that express their respective invasiveness signature at a high level (Figure 5B, yellow arrows). Again, the number of invasive-like DCIS tumors that we identify is consistent with previous reports indicating that only a small proportion of pre-invasive tumors have the capacity to become invasive over time.

DISCUSSION

Today, approximately one in five breast malignancies detected by mammography are diagnosed as DCIS. DCIS continues to pose a difficult clinical challenge as we still cannot distinguish cases that would remain indolent and therefore do not require intensive treatment from cases that are likely to progress to IDC. Therefore almost all DCIS cases are treated with excision, often followed by radiation and hormonal therapy (Allred et al., 2012; Punglia et al., 2013; Wärnberg et al., 2014). However, at least 50% of DCIS would never progress to an invasive state if left untreated, implying many women needlessly undergo treatment with potentially harmful side-effects (Vatovec et al., 2014). An increased understanding of the tumor properties that drive disease progression, and an investigation into whether some of these properties are already present within (a minority of) pre-invasive lesions, is needed.

Our finding that gene expression profiles cluster patients primarily by breast cancer subtype and then according to disease stage is concordant with previous reports. Moreover, we demonstrate that signatures derived in unstratified analysis to differentiate DCIS from IDC are indeed confounded by intrinsic subtype; the observed differences in expression are more associated with subtype than with stage of disease. We note that previous studies may have been unable to observe these confounding effects because of their smaller sample size.

In order to overcome the association between intrinsic subtype and classification markers of invasiveness, we stratified tumors according to their intrinsic subtype and identified tumor properties that differentiate DCIS from IDC. Although our data set is relatively large, we note that these stratified analyses reduced the power to identify differentially expressed genes, particularly within the NormL subtype where no genes were identified. Nonetheless, these analyses highlighted substantial differences between the subtypes, in turn leading to the hypothesis that each subtype undergoes a distinct evolutionary course of disease progression from pre-invasive to invasive stage. However, among the different subtypes, it is interesting that most of these properties do not represent molecular

Figure 2. Subtype-Stratified Classification Markers of Disease Stage

(A) Genes that are differentially expressed between DCIS and IDC in either an unstratified or PAM50-stratified analysis are displayed as heatmaps within each subtype. The patients are ordered according to an increasing rank-sum by their intrinsic subtype and type of lesion. The gene tick marks to the left of the heatmaps are colored according to whether they are significant in each analysis.

(B) Venn diagram representing number of genes differentially expressed between DCIS and IDC within each intrinsic subtype.

(C) Venn diagram representing number of miRNA differentially expressed between DCIS and IDC within each intrinsic subtype.

Figure 3. Patient-Signature Heatmap of Signatures of Progression in Cancer

(A) Subtype-specific performance of signatures is represented. The colors are proportional to the rank of the classifier within the specific patient cohort, with red representing the highest-performing signatures. The ticks represent the level of significance of the classifier by Fisher's exact test. The percentage of significant signatures at p value < 0.05 is shown in parentheses.

(B) Yellow and purple shaded columns correspond to DCIS and IDC, respectively. The signatures (rows) are ordered by their ability to predict tissue type in the unstratified cohort. The dark and light shades correspond to correct and incorrect predictions, respectively. The patients (columns) are ordered by the degree of agreement of predictions across all signatures. A description of the signatures used is available in Table S1.

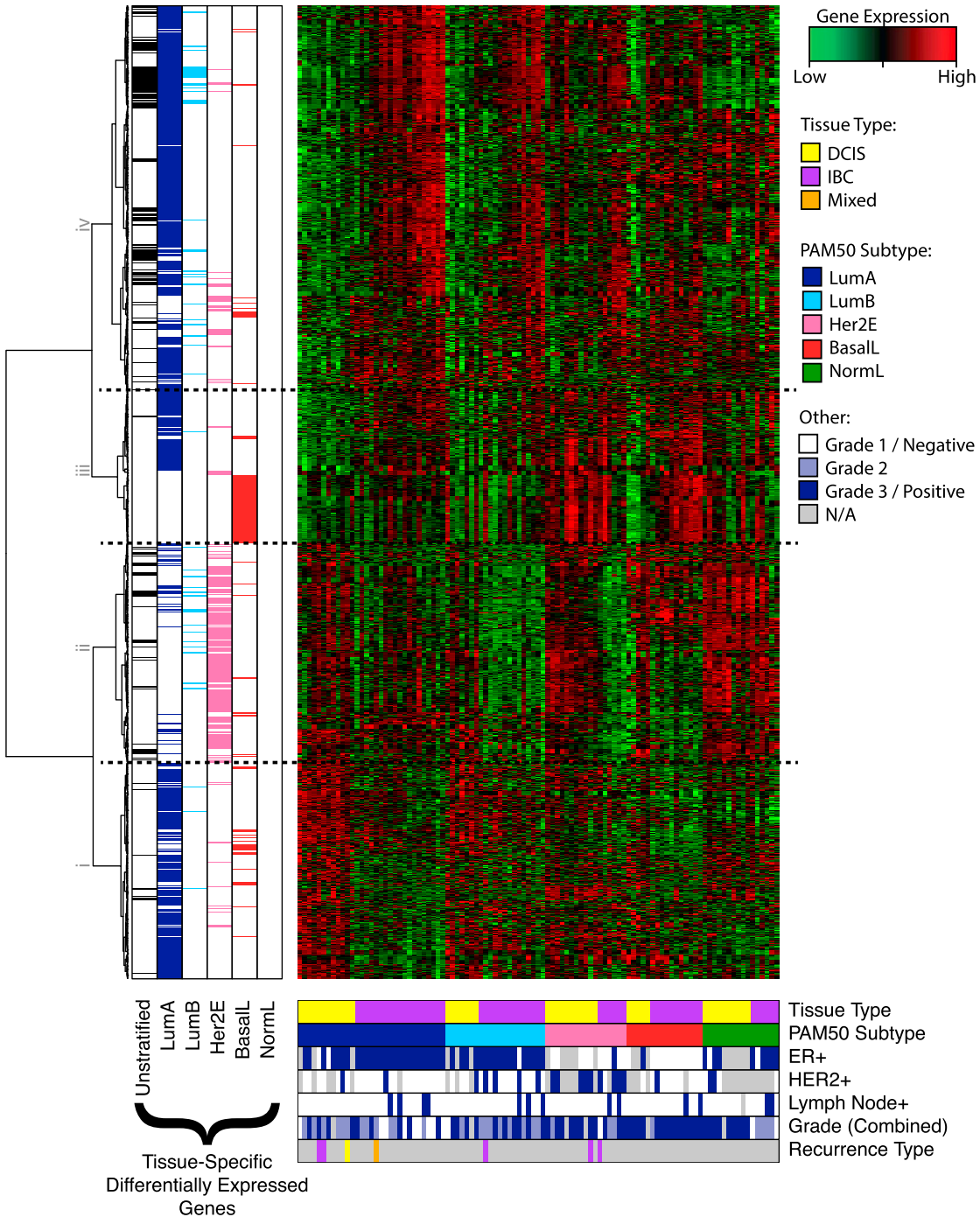
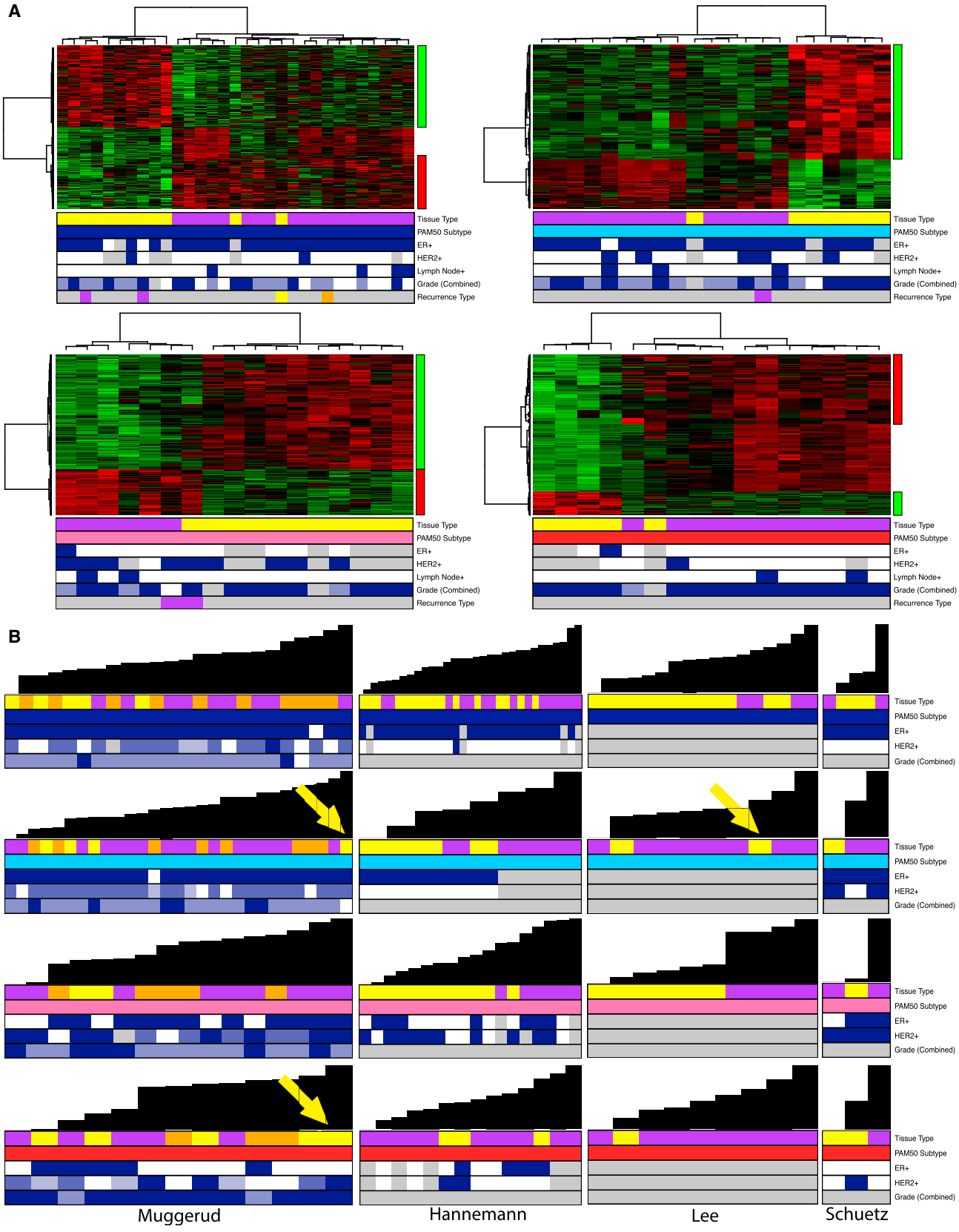


Figure 4. Subtype-Stratified Classification Markers of Tumor Stage Hierarchically Clustered across All Samples

Heatmap of genes that are differentially expressed between DCIS and IDC in either an unstratified or subtype-stratified analysis. The patients are sorted by their intrinsic subtype and type of lesion. The gene tick marks to the left of the heatmap are colored according to whether they are significant in each analysis.

events that occur within the epithelial tumor cells, but rather reflect changes that include involvement of the microenvironment. This supports the increasing evidence of the role that non-epithelial events play in tumor progression and disease outcome.

Although we identified miRNAs as differentially expressed between DCIS and IDC within each subtype, none were shared between the subtypes. Some of these have been previously implicated in cancer. For example, the miR-10 family has been demonstrated to be involved in Hox gene transcription factor



(legend on next page)

signaling, with high expression associated with various cancer types, including breast (Lund, 2010; Ma et al., 2007). miR-298 may be involved in chemoresistance in breast cancer (Bao et al., 2012). miR-34c-5p is induced by a hypoxia (Xu et al., 2012), and miR-192 is involved in p53-mediated MDM2 expression (Pichiorri et al., 2010). miR-133b and miR-136 are involved in apoptosis (Patron et al., 2012; Yang et al., 2012). Additionally, miR-133b regulates the oncogene MET (Hu et al., 2010) and promotes cell proliferation through ERK and AKT1 (Qin et al., 2012). Interestingly, miR-K12-5* is derived from Kaposi's sarcoma-associated herpesvirus, which is sometimes present in the lymphocytes of individuals. Its high expression in invasive BasalL breast cancer is consistent with the high expression of lymphocytic markers in these tumors. We note, however, that as fewer DCIS and IDC samples were profiled for miRNA expression, this analysis had limited power to detect differentially expressed miRNA.

At the DNA level, tumors displayed heterogeneous profiles, even after stratification by intrinsic subtype. This is in line with previous studies of the DNA profiles of DCIS (Buerger et al., 1999a, 1999b; O'Connell et al., 1998), which have suggested that they are only partially correlated with gene expression profiles derived from the same tumors. Similar mutation patterns and frequencies in *TP53* and *PIK3CA* were found in DCIS lesions compared with invasive tumors, which is consistent with what has previously been reported (Miron et al., 2010). Mutations in these two genes, which are the most frequently mutated in breast cancer, are strongly correlated with breast cancer subtypes and are therefore less likely to reflect within subtype differences between DCIS and IDC. Future studies based on massively parallel sequencing may identify clonal differences between DCIS and IDC and aid in the identification of somatic mutations and chromosomal aberrations that differentiate between disease stages (Kaur et al., 2013). Although our study was conducted on a relatively large sample set of DCIS, having a higher number of pre-invasive lesions would improve the power to extend our findings in the subtype-stratified analyses.

A critical question that remains is whether it is possible to predict which DCIS lesions are most likely to progress if left untreated, in order to spare patients from unnecessary treatment. Although previous studies and tests have sought to identify markers of disease recurrence in patients with DCIS (Rakovitch et al., 2015; Solin et al., 2013), it is important to differentiate between an in situ recurrence and an invasive recurrence. To this end, we used a series of leave-one-out analyses to identify a minority of DCIS tumors that have gene expression profiles resembling invasive tumors. Although all DCIS cases were assessed by pathologists to ensure that only pure DCIS were included in this group, it is possible that, due to spatial heterogeneity within the lesions, our analysis identifies tissues with micro-invasive components that were not otherwise reported. While studies using microdissection would resolve this concern, they may lack the sensitivity to identify such potentially important cases. We additionally used the identification of these lesions to further develop subtype-specific signatures that predict invasiveness. Interestingly, all of these signatures again highlight a role for the tumor microenvironment in driving disease invasion. In particular, gene enrichment for immune response pathways was observed in multiple subtypes. Together, these results suggest that such properties could be used to differentiate between DCIS patients who require further treatment or who could be spared disease therapy.

EXPERIMENTAL PROCEDURES

This study has been approved by the appropriate ethical committees (Norwegian Regional Committee for Medical Research Ethics, 1.2006.1607, and the Ethics Committee at Uppsala University, Sweden, Dnr 2005:118).

Tissue Processing

Tumor tissues were obtained from the Fresh Tissue Biobank, Department of Pathology, Uppsala University Hospital, Sweden; the Breast Cancer Tissue Bank, St. Olav's University Hospital, Trondheim, Norway; the Norwegian Radium Hospital and Ullevål University Hospital, Oslo University Hospital, Oslo, Norway; and Akershus University Hospital, Lørenskog, Norway. Expert pathologists reviewed all cases to ensure that no invasive components were found in the pure DCIS. For grade, DCIS lesions were classified according to European Organization for Research and Treatment of Cancer (EORTC) scores, and IDC and the invasive part of tumors with a mixed diagnosis were classified by using the Elston-Ellis scores. Individual tissue clinical characteristics are found in Table S1 and a summary of these characteristics is available in Table 1.

Microarray Analysis

mRNA and miRNA expression analysis was performed using Agilent SurePrint G3 Human GE 8x60K Microarrays and Human miRNA Microarray Release 16.0, 8x60K, respectively. Copy-number analysis was performed using Affymetrix Genome-Wide Human SNP array 6.0. Details on data analysis can be found in the Supplemental Information.

Statistical Methods Rank-Sum Algorithm for Tumor Ordering

We use the rank-sum linear ordering algorithm to sort tumors according to increasing activation levels of signatures. For a given signature with k genes, each gene univariately ranks all patients according to the level of expression of that gene and that patient. For many signatures in the literature including the majority of signatures used here, we have prior knowledge regarding whether each of the genes is over- or under-expressed in the control and experimental groups. For example, if we had a signature of ER activation, we might expect that ESR1 would be overexpressed in the ER-positive subgroup of patients. Symmetrically, we might expect that markers of the BasalL subtype might be under-expressed. For a gene that is expected to be overexpressed in our experiment, each patient is ranked from 1 to m where 1 is the least amount of expression witnessed for that gene, and m is the highest observed expression of that gene. Similarly, for genes expected to be under-expressed, the patients are ranked from m to 1. The sum of ranks for

Figure 5. Subtype-Specific Signatures of Tumor Progression

(A) Heatmaps showing differentially expressed genes between DCIS and IDC-like lesions within each intrinsic subtype. A total of five DCIS were classified as invasive-like and cluster here among IDC. The genes that follow similar patterns within true IDC and invasive-like DCIS are shown by red and green bars to the right of heatmaps and form the basis of our progression signatures.

(B) Progression signatures were applied to four other data sets (Hannemann et al., 2006; Lee et al., 2012; Muggerud et al., 2010; Schuetz et al., 2006). Here, patients are each ordered from left to right according to a rank-sum algorithm (see Supplemental Information). A small number of DCIS in these data sets order to the far right, indicating that they possess invasive-like components within their genomic profiles.

Table 1. Summary of Patient and Tumor Characteristics

	DCIS	IDC	Mixed	Total
Number of samples	59	85	16	160
Number of expression arrays	46	56	0	102
Number of microRNA arrays	26	14	0	40
Number of SNP arrays	42	67	16	125
Age in years, median (range)	56.5 (30–82)	57 (36–90)	57 (27–89)	57 (27–90)
Size in millimeters, median (range)	25 (7–60)	11 (1–30)	15 (8–40)	13 (1–60)
Elston grade, I/II/III	N/A	25/32/28	6/7/3	31/39/31
EORTC grade, I/II/III	3/14/35	N/A	0/6/7	3/20/42
Combined grade, I/II/III	3/15/35	26/32/27	6/7/3	35/54/65
ER+/ER− (%)	26/11 (70)	64/21 (75)	9/7 (56)	99/39 (72)
PR+/PR− (%)	23/14 (62)	59/26 (69)	8/8 (50)	90/48 (65)
HER2+/HER2− (%)	12/22 (35)	23/53 (30)	13/2 (87)	48/77 (38)
Lymph node+/lymph node− (%)	0/46 (0)	13/43 (23)	0/0 (N/A)	13/88 (13)
LumA (%)	17 (29)	28 (33)	4 (25)	49 (31)
LumB (%)	10 (17)	22 (26)	1 (6)	33 (21)
Her2E (%)	12 (20)	11 (13)	5 (31)	28 (18)
BasalL (%)	8 (14)	14 (17)	3 (19)	25 (16)
NormL (%)	12 (20)	9 (11)	3 (19)	24 (15)

Combined grade uses Elston grade in invasive and mixed tissues and EORTC grade in DCIS tissues; variables have some missing data; and percentages are calculated with the missing data omitted.

each patient are calculated, and the patients are ordered left to right (from least to greatest sum), representing least to greatest activation of the signature accordingly.

Methodology to compute statistical associations between patient orderings and the state of clinicopathological variables followed our previously described approach. The Mann-Whitney-Wilcoxon test was used to determine associations with type of lesion, whereas the Kruskal-Wallis test was used to determine association with intrinsic subtype.

Identification of Invasive-like DCIS Tumors

Our method to identify DCIS tumors with invasive-like molecular components to their profiles proceeded as follows. (1) Within each subtype, DCIS tumors were sequentially removed under a leave-one-out approach, and differentially expressed genes were identified between the remaining DCIS versus IDC samples as previously described. All samples belonging to that subtype (included the sample that had been removed) were then hierarchically clustered. In some cases, the left-out DCIS clustered most closely with IDC. (2) We collected a set of over 6,421 gene signatures from MSigDB (Subramanian et al., 2005), GeneSigDB (Culhane et al., 2012), and the literature. For each signature and within each subtype, samples were sorted by our rank-sum algorithm for tumor ordering (described above). Within each subtype, DCIS tumors were sequentially removed under a leave-one-out approach, and gene signatures with significantly different sample ranks were identified between the remaining DCIS versus IDC samples as previously described. All samples belonging to that subtype (included the sample that had been removed) were then hierarchically clustered over the sample ranks of filtered signatures. Those DCIS that clustered among IDC in both (1) and (2) were designated as having invasive-like properties.

After we identified a small number of invasive-like DCIS tumors belonging to the LumA, LumB, Her2E, and BasalL subtypes, we re-classified these tumors into the IDC group. For each of these four subtypes, the genes differentially expressed between DCIS and IDC were identified as previously described. These genes were next hierarchically clustered and plotted into a heatmap. Gene clusters that followed consistent patterns within IDC and invasive-like DCIS were visually determined and formed the basis for each progression signature.

ACCESSION NUMBERS

The accession number for the microarray data reported in this paper is GEO: GSE59248. The accession number for the Raw Affymetrix Array SNP 6.0 data reported in this paper is European Genome-phenome Archive (EGA): EGAD00010000942.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.06.051.

CONSORTIA

The members of OSBREAC are Torill Sauer, Jürgen Geisler, Solveig Hofvind, Elin Borgen, Anne-Lise Børresen-Dale, Olav Engebråten, Øystein Fodstad, Øystein Garred, Gry Aarum Geitvik, Rolf Kåresen, Bjørn Naume, Gunhild Mari Mælandsmo, Hege G. Russnes, Ellen Schlichting, Therese Sørlie, Ole Christian Lingjærde, Vessela Kristensen, Kristine Kleivi Sahlberg, Helle Kristine Skjerven, and Britt Fritzman.

AUTHOR CONTRIBUTIONS

Conceptualization, R.L., F.W., M.H., and T.S.; Methodology, R.L., M.H., and T.S.; Software, R.L. and M.H.; Formal Analysis, R.L. and V.V.; Investigation; R.L., M.R.A., H.H.M., M.H., and T.S.; Resources, OSBREAC, S.L., F.W., V.K., and A.-L.B.-D.; Funding Acquisition, M.H., A.-L.B.-D., F.W., and T.S.; Writing, R.L., M.H., and T.S.

ACKNOWLEDGMENTS

We thank Beathe Sitter for initiating the sample collection at St. Olav's University Hospital. We also thank other members of the Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, and the M.H. lab for their insightful discussions. This research was supported by funds

from Helse Sør-Øst (2012056), the Norwegian Cancer Society (420056), and the Oslo University Hospital (to T.S.); the K.G. Jebsen Centre for Breast Cancer Research (to A.-L.B.-D.); the United States Department of Defense (W81XWH-10-1-0299) and the McGill University Faculty of Medicine (to R.L.); and the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada (to M.H.).

Received: November 30, 2015
Revised: May 3, 2016
Accepted: June 10, 2016
Published: July 7, 2016

REFERENCES

Allred, D.C., Wu, Y., Mao, S., Nagtegaal, I.D., Lee, S., Perou, C.M., Mohsin, S.K., O'Connell, P., Tsimelzon, A., and Medina, D. (2008). Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. Clin. Cancer Res. *14*, 370–378.

Allred, D.C., Anderson, S.J., Paik, S., Wickerham, D.L., Nagtegaal, I.D., Swain, S.M., Mamounas, E.P., Julian, T.B., Geyer, C.E., Jr., Costantino, J.P., et al. (2012). Adjuvant tamoxifen reduces subsequent breast cancer in women with estrogen receptor-positive ductal carcinoma in situ: a study based on NSABP protocol B-24. J. Clin. Oncol. *30*, 1268–1273.

Bao, L., Hazari, S., Mehra, S., Kaushal, D., Moroz, K., and Dash, S. (2012). Increased expression of P-glycoprotein and doxorubicin chemoresistance of metastatic breast cancer is regulated by miR-298. Am. J. Pathol. *180*, 2490–2503.

Berman, H., Zhang, J., Crawford, Y.G., Gauthier, M.L., Fordyce, C.A., McDermott, K.M., Sigaroudinia, M., Kozakiewicz, K., and Tlsty, T.D. (2005). Genetic and epigenetic changes in mammary epithelial cells identify a subpopulation of cells involved in early carcinogenesis. Cold Spring Harb. Symp. Quant. Biol. *70*, 317–327.

Buerger, H., Otterbach, F., Simon, R., Poremba, C., Diallo, R., Decker, T., Riethdorf, L., Brinkschmidt, C., Dockhorn-Dworniczak, B., and Boecker, W. (1999a). Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. J. Pathol. *187*, 396–402.

Buerger, H., Otterbach, F., Simon, R., Schäfer, K.L., Poremba, C., Diallo, R., Brinkschmidt, C., Dockhorn-Dworniczak, B., and Boecker, W. (1999b). Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes. J. Pathol. *189*, 521–526.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Chin, K., de Solorzano, C.O., Knowles, D., Jones, A., Chou, W., Rodriguez, E.G., Kuo, W.L., Ljung, B.M., Chew, K., Myambo, K., et al. (2004). In situ analyses of genome instability in breast cancer. Nat. Genet. *36*, 984–988.

Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-Morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., van de Wiel, M.A., Green, A.R., et al. (2007). High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. Genome Biol. *8*, R215.

Clark, S.E., Warwick, J., Carpenter, R., Bowen, R.L., Duffy, S.W., and Jones, J.L. (2011). Molecular subtyping of DCIS: heterogeneity of breast cancer reflected in pre-invasive disease. Br. J. Cancer *104*, 120–127.

Cowell, C.F., Weigelt, B., Sakr, R.A., Ng, C.K., Hicks, J., King, T.A., and Reis-Filho, J.S. (2013). Progression from ductal carcinoma in situ to invasive breast cancer: revisited. Mol. Oncol. *7*, 859–869.

Culhane, A.C., Schröder, M.S., Sultana, R., Picard, S.C., Martinelli, E.N., Kelly, C., Haibe-Kains, B., Kapushesky, M., St Pierre, A.A., Flahive, W., et al. (2012). GeneSigDB: a manually curated database and resource for analysis of gene expression signatures. Nucleic Acids Res. *40*, D1060–D1066.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352.

Ernster, V.L., Ballard-Barbash, R., Barlow, W.E., Zheng, Y., Weaver, D.L., Cutter, G., Yankaskas, B.C., Rosenberg, R., Carney, P.A., Kerlikowske, K., et al. (2002). Detection of ductal carcinoma in situ in women undergoing screening mammography. J. Natl. Cancer Inst. *94*, 1546–1554.

Espina, V., Mariani, B.D., Gallagher, R.I., Tran, K., Banks, S., Wiedemann, J., Huryk, H., Mueller, C., Adamo, L., Deng, J., et al. (2010). Malignant precursor cells pre-exist in human breast DCIS and require autophagy for survival. PLoS ONE *5*, e10240.

Falk, R.S., Hofvind, S., Skaane, P., and Haldorsen, T. (2011). Second events following ductal carcinoma in situ of the breast: a register-based cohort study. Breast Cancer Res. Treat. *129*, 929–938.

Fazzari, M.J., and Greally, J.M. (2004). Epigenomics: beyond CpG islands. Nat. Rev. Genet. *5*, 446–455.

Fitzgibbons, P.L., Henson, D.E., and Hutter, R.V.; Cancer Committee of the College of American Pathologists (1998). Benign breast changes and the risk for subsequent breast cancer: an update of the 1985 consensus statement. Arch. Pathol. Lab. Med. *122*, 1053–1055.

Fleischer, T., Frigessi, A., Johnson, K.C., Edvardsen, H., Touleimat, N., Klajic, J., Riis, M.L., Haakensen, V.D., Wärnberg, F., Naume, B., et al. (2014). Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. Genome Biol. *15*, 435.

Hannemann, J., Velds, A., Halfwerk, J.B., Kreike, B., Peterse, J.L., and van de Vijver, M.J. (2006). Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res. *8*, R61.

Hu, M., Yao, J., Carroll, D.K., Weremowicz, S., Chen, H., Carrasco, D., Richardson, A., Violette, S., Nikolskaya, T., Nikolsky, Y., et al. (2008). Regulation of in situ to invasive breast carcinoma transition. Cancer Cell *13*, 394–406.

Hu, G., Chen, D., Li, X., Yang, K., Wang, H., and Wu, W. (2010). miR-133b regulates the MET proto-oncogene and inhibits the growth of colorectal cancer cells in vitro and in vivo. Cancer Biol. Ther. *10*, 190–197.

Hughes, L.L., Wang, M., Page, D.L., Gray, R., Solin, L.J., Davidson, N.E., Lowen, M.A., Ingle, J.N., Recht, A., and Wood, W.C. (2009). Local excision alone without irradiation for ductal carcinoma in situ of the breast: a trial of the Eastern Cooperative Oncology Group. J. Clin. Oncol. *27*, 5319–5324.

Jang, M., Kim, E., Choi, Y., Lee, H., Kim, Y., Kim, J., Kang, E., Kim, S.W., Kim, I., and Park, S. (2012). FGFR1 is amplified during the progression of in situ to invasive breast carcinoma. Breast Cancer Res. *14*, R115.

Kaur, H., Mao, S., Shah, S., Gorski, D.H., Krawetz, S.A., Sloane, B.F., and Mattingly, R.R. (2013). Next-generation sequencing: a powerful tool for the discovery of molecular markers in breast ductal carcinoma in situ. Expert Rev. Mol. Diagn. *13*, 151–165.

Knudsen, E.S., Ertel, A., Davicioni, E., Kline, J., Schwartz, G.F., and Witkiewicz, A.K. (2012). Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. Breast Cancer Res. Treat. *133*, 1009–1024.

Lee, S., Stewart, S., Nagtegaal, I., Luo, J., Wu, Y., Colditz, G., Medina, D., and Allred, D.C. (2012). Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. Cancer Res. *72*, 4574–4586.

Livasy, C.A., Perou, C.M., Karaca, G., Cowan, D.W., Maia, D., Jackson, S., Tse, C.K., Nyante, S., and Millikan, R.C. (2007). Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum. Pathol. *38*, 197–204.

Lund, A.H. (2010). miR-10 in development and cancer. Cell Death Differ. *17*, 209–214.

Ma, X.J., Salunga, R., Tuggle, J.T., Gaudet, J., Enright, E., McQuary, P., Payette, T., Pistone, M., Stecker, K., Zhang, B.M., et al. (2003). Gene expression profiles of human breast cancer progression. Proc. Natl. Acad. Sci. USA *100*, 5974–5979.

Ma, L., Teruya-Feldstein, J., and Weinberg, R.A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature *449*, 682–688.

Ma, X.J., Dahiya, S., Richardson, E., Erlander, M., and Sgroi, D.C. (2009). Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. *11*, R7.

Miron, A., Varadi, M., Carrasco, D., Li, H., Luongo, L., Kim, H.J., Park, S.Y., Cho, E.Y., Lewis, G., Kehoe, S., et al. (2010). PIK3CA mutations in in situ and invasive breast carcinomas. Cancer Res. *70*, 5674–5678.

Muggerud, A.A., Hallett, M., Johnsen, H., Kleivi, K., Zhou, W., Tahmasebpoor, S., Amini, R.M., Botling, J., Børresen-Dale, A.L., Sørlie, T., and Wärnberg, F. (2010). Molecular diversity in ductal carcinoma in situ (DCIS) and early invasive breast cancer. Mol. Oncol. *4*, 357–368.

Narod, S.A., Iqbal, J., Giannakeas, V., Sopik, V., and Sun, P. (2015). Breast cancer mortality after a diagnosis of ductal carcinoma in situ. JAMA Oncol. *1*, 888–896.

Newburger, D.E., Kashef-Haghighi, D., Weng, Z., Salari, R., Sweeney, R.T., Brunner, A.L., Zhu, S.X., Guo, X., Varma, S., Troxell, M.L., et al. (2013). Genome evolution during progression to breast cancer. Genome Res. *23*, 1097–1108.

O'Connell, P., Pekkel, V., Fuqua, S.A., Osborne, C.K., Clark, G.M., and Allred, D.C. (1998). Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. J. Natl. Cancer Inst. *90*, 697–703.

Page, D.L., Dupont, W.D., Rogers, L.W., and Landenberger, M. (1982). Intraductal carcinoma of the breast: follow-up after biopsy only. Cancer *49*, 751–758.

Page, D.L., Dupont, W.D., Rogers, L.W., Jensen, R.A., and Schuyler, P.A. (1995). Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer *76*, 1197–1200.

Park, K., Han, S., Kim, H.J., Kim, J., and Shin, E. (2006). HER2 status in pure ductal carcinoma in situ and in the intraductal and invasive components of invasive ductal carcinoma determined by fluorescence in situ hybridization and immunohistochemistry. Histopathology *48*, 702–707.

Patron, J.P., Fendler, A., Bild, M., Jung, U., Müller, H., Arntzen, M.O., Piso, C., Stephan, C., Thiede, B., Mollenkopf, H.J., et al. (2012). MiR-133b targets antiapoptotic genes and enhances death receptor-induced apoptosis. PLoS ONE *7*, e35345.

Pichiorri, F., Suh, S.S., Rocci, A., De Luca, L., Taccioli, C., Santhanam, R., Zhou, W., Benson, D.M., Jr., Hofmainster, C., Alder, H., et al. (2010). Downregulation of p53-inducible microRNAs 192, 194, and 215 impairs the p53/MDM2 autoregulatory loop in multiple myeloma development. Cancer Cell *18*, 367–381.

Porter, D., Lahti-Domenici, J., Keshaviah, A., Bae, Y.K., Argani, P., Marks, J., Richardson, A., Cooper, A., Strausberg, R., Riggins, G.J., et al. (2003). Molecular markers in ductal carcinoma in situ of the breast. Mol. Cancer Res. *1*, 362–375.

Punglia, R.S., Schnitt, S.J., and Weeks, J.C. (2013). Treatment of ductal carcinoma in situ after excision: would a prophylactic paradigm be more appropriate? J. Natl. Cancer Inst. *105*, 1527–1533.

Qin, W., Dong, P., Ma, C., Mitchelson, K., Deng, T., Zhang, L., Sun, Y., Feng, X., Ding, Y., Lu, X., et al. (2012). MicroRNA-133b is a key promoter of cervical carcinoma development through the activation of the ERK and AKT1 pathways. Oncogene *31*, 4067–4075.

Rakovitch, E., Nofech-Mozes, S., Hanna, W., Baehner, F.L., Saskin, R., Butler, S.M., Tuck, A., Sengupta, S., Elavathil, L., Jani, P.A., et al. (2015). A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. Breast Cancer Res. Treat. *152*, 389–398.

Robanus-Maandag, E.C., Bosch, C.A., Kristel, P.M., Hart, A.A., Faneyte, I.F., Nederlof, P.M., Peterse, J.L., and van de Vijver, M.J. (2003). Association of C-MYC amplification with progression from the in situ to the invasive stage in C-MYC-amplified breast carcinomas. J. Pathol. *201*, 75–82.

Sanders, M.E., Schuyler, P.A., Dupont, W.D., and Page, D.L. (2005). The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer *103*, 2481–2484.

Schuetz, C.S., Bonin, M., Clare, S.E., Nieselt, K., Sotlar, K., Walter, M., Fehm, T., Solomayer, E., Riess, O., Wallwiener, D., et al. (2006). Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res. *66*, 5278–5286.

Solin, L.J., Gray, R., Baehner, F.L., Butler, S.M., Hughes, L.L., Yoshizawa, C., Cherbavaz, D.B., Shak, S., Page, D.L., Sledge, G.W., Jr., et al. (2013). A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. J. Natl. Cancer Inst. *105*, 701–710.

Stretch, C., Khan, S., Asgarian, N., Eisner, R., Vaisipour, S., Damaraju, S., Graham, K., Bathe, O.F., Steed, H., Greiner, R., and Baracos, V.E. (2013). Effects of sample size on differential gene expression, rank order and prediction accuracy of a gene signature. PLoS ONE *8*, e65380.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

van de Vijver, M.J., Peterse, J.L., Mooi, W.J., Wisman, P., Lomans, J., Dalesio, O., and Nusse, R. (1988). Neu-protein overexpression in breast cancer. Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. N. Engl. J. Med. *319*, 1239–1245.

Vargas, A.C., McCart Reed, A.E., Waddell, N., Lane, A., Reid, L.E., Smart, C.E., Cocciardi, S., da Silva, L., Song, S., Chenevix-Trench, G., et al. (2012). Gene expression profiling of tumour epithelial and stromal compartments during breast cancer progression. Breast Cancer Res. Treat. *135*, 153–165.

Vatovec, C., Erten, M.Z., Kolodinsky, J., Brown, P., Wood, M., James, T., and Sprague, B.L. (2014). Ductal carcinoma in situ: a brief review of treatment variation and impacts on patients and society. Crit. Rev. Eukaryot. Gene Expr. *24*, 281–286.

Vincent-Salomon, A., Lucchesi, C., Gruel, N., Raynal, V., Pierron, G., Goudefroye, R., Reyal, F., Radvanyi, F., Salmon, R., Thiery, J.P., et al.; breast cancer study group of the Institut Curie (2008). Integrated genomic and transcriptomic analysis of ductal carcinoma in situ of the breast. Clin. Cancer Res. *14*, 1956–1965.

Volinia, S., Galasso, M., Sana, M.E., Wise, T.F., Palatini, J., Huebner, K., and Croce, C.M. (2012). Breast cancer signatures for invasiveness and prognosis defined by deep sequencing of microRNA. Proc. Natl. Acad. Sci. USA *109*, 3024–3029.

Wärnberg, F., Garmo, H., Emdin, S., Hedberg, V., Adwall, L., Sandelin, K., Ringberg, A., Karlsson, P., Arnesson, L.G., Anderson, H., et al. (2014). Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS Trial. J. Clin. Oncol. *32*, 3613–3618.

Widschwendter, M., and Jones, P.A. (2002). DNA methylation and breast carcinogenesis. Oncogene *21*, 5462–5482.

Xu, X., Wang, S., Liu, J., Dou, D., Liu, L., Chen, Z., Ye, L., Liu, H., He, Q., Raj, J.U., and Gao, Y. (2012). Hypoxia induces downregulation of soluble guanylyl cyclase β1 by miR-34c-5p. J. Cell Sci. *125*, 6117–6126.

Yang, Y., Wu, J., Guan, H., Cai, J., Fang, L., Li, J., and Li, M. (2012). MiR-136 promotes apoptosis of glioma cells by targeting AEG-1 and Bcl-2. FEBS Lett. *586*, 3608–3612.