Published in final edited form as:

Genes Chromosomes Cancer. 2012 December ; 51(12): 1067–1078. doi:10.1002/gcc.21991.

Differential Copy Number Aberrations in Novel Candidate Genes Associated with Progression from *In Situ* to Invasive Ductal Carcinoma of the Breast

Shaoxi Liao¹, Mohamed M. Desouki², Daniel P. Gaile³, Lori Shepherd⁴, Norma J. Nowak^{5,6}, Jeffrey Conroy⁵, William T. Barry⁷, and Joseph Geradts¹

¹Department of Pathology, Duke University Medical Center, Durham, NC

²Department of Pathology, Vanderbilt University Medical Center, Nashville, TN

³Department of Biostatistics, State University of New York, Buffalo, NY

⁴Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY

⁵Department of Cancer Genetics, Roswell Park Cancer Institute, Buffalo, NY

⁶Department of Biochemistry, State University of New York, Buffalo, NY

⁷Department of Biostatistics, Duke University Medical Center, Durham, NC

Abstract

Only a minority of intraductal carcinomas of the breast give rise to stromally invasive disease. We microdissected 206 paraffin blocks representing 116 different cases of low grade DCIS. Fifty-five were pure DCIS cases without progression to invasive carcinoma. Sixty-one cases had a small invasive component. DNA was extracted from microdissected sections and hybridized to high density BAC arrays. Array comparative genomic hybridization (aCGH) analysis of 118 hybridized DNA samples yielded data on 69 samples that were suitable for further statistical analysis. This cohort included 20 pure DCIS cases (PD), 25 mixed DCIS (MD), and 24 mixed invasive carcinoma samples (MI). Pure DCIS cases had a higher frequency of DNA copy number changes than mixed DCIS cases, and the latter had similar DNA profiles compared to paired invasive carcinomas. Copy number changes on 13 chromosomal arms occurred at different rates in PD *versus* MD lesions. Eight of 19 candidate genes residing at those loci were confirmed to have differential copy number changes by quantitative PCR. *NCOR2/SMRT* and *NR4A1* (both on 12q), *DYNLRB2* (16q), *CELSR1*, *UPK3A* and *ST13* (all on 22q) were more frequently amplified in PD. Moreover, *NCOR2*, *NR4A1* and *DYNLRB2* showed more frequent copy number losses in MD. *GRAP2* (22q) was more often amplified in MD, whereas *TAF1C* (16q) was more commonly deleted in PD. A multi-gene model comprising these candidate genes discriminated between PD and MD lesions with high accuracy. These findings suggest that the propensity to invade the stroma may be encoded in the genome of intraductal carcinomas.

Introduction

Ductal carcinoma *in situ* (DCIS) is considered a non-obligate precursor of invasive breast cancer. Only a minority of those lesions are thought to have the capacity to progress to invasive carcinoma (Eusebi et al. 1994; Frykberg and Bland 1994). Currently, there are no generally accepted features that allow us to reliably distinguish those DCIS lesions that are

likely to progress and the larger number of intraductal tumors that do not have this capacity. Consequently, many patients with DCIS receive unnecessary surgical and adjuvant treatment. Prediction of progression is particularly challenging for low grade (LG) DCIS. It has been suggested that less than 30% of these neoplasms will progress to invasive carcinoma within 20 years (Sanders et al. 2005).

The transition from *in situ* to invasive carcinoma is a critical step in mammary tumorigenesis. In essence, DCIS is a benign disease that cannot spread outside the breast since the neoplastic cells do not have access to lymphatic or vascular channels. In recent years it has become evident that the myoepithelial cells and the stroma surrounding the ducts and ductules bidirectionally interact with the *in situ* carcinoma (Polyak and Hu 2005; Bombonati and Sgroi 2011). However, it remains unknown which molecular pathologic events tip the balance in favor of invasion. We hypothesized that alterations in the genome of the intraductal tumor cells play a major role in determining whether a particular epithelial neoplasm will interact with its microenvironment in a way that enables the tumor cells to break through the basement membrane and invade the stroma. Specifically, in this study we wanted to test the hypothesis that certain DNA copy number gains and losses in the intraductal tumor cells could be linked to the presence of stromal invasion. Our primary objective was not to identify genes driving stromal invasion, but rather to uncover candidate genes that may be used as prognostic biomarkers. Our approach included whole genome profiling of highly enriched neoplastic cell populations microdissected from LGDCIS lesions with and without associated invasive carcinoma using high density comparative genomic hybridization (CGH) microarrays, identifying differential copy number gains or losses in indolent (without evidence of invasion) *versus* aggressive neoplasms (that had progressed to invasive carcinoma), and validating copy number changes in candidate genes by quantitative PCR (qPCR). Our experiments revealed a DNA signature that indeed correlated with the invasive potential of LGDCIS.

Materials and Methods

Tissue samples and microdissection

This study was approved by the Institutional Review Boards of Roswell Park Cancer Institute (RPCI) and Duke University Medical Center (DUMC). From the surgical pathology files at RPCI and DUMC, we retrieved 116 cases of formalin-fixed and paraffin-embedded (FFPE) LGDCIS. We followed the Van Nuys grading system for DCIS, in which low grade DCIS is defined as DCIS of any architectural subtype with a nuclear grade of 1 or 2, without prominent centriductal necrosis (Silverstein et al. 1996). The predominant growth patterns were solid, cribriform and papillary. Lesions with a nuclear grade of 3 or prominent necrosis were excluded. All cases were reviewed by the two study pathologists (MMD and JG). Fifty-five cases had no associated invasive component and showed no evidence of recurrence or progression to invasive disease within 5 years prior to sample processing (PD = pure DCIS). The PD cases included in the final analysis (n=20) were resected 5–11 years prior to DNA extraction. The majority of PD cases underwent lumpectomy (+/- reexcision), and most of these cases received postoperative radiation therapy. The clinical follow up was recently updated for the Duke cases. None of the 11 cases developed invasive carcinoma in the ipsilateral breast (median follow up 77 months); three patients died of unrelated causes. It is acknowledged that LGDCIS may recur after long periods of time, and we cannot rule out that recurrences may develop in the PD group with longer follow up, although overall the long-term recurrence rate for properly excised LGDCIS is less than 5%. Sixty-one cases contained a small (< 2 cm) component of invasive carcinoma (MD = mixed DCIS, MI = mixed invasive carcinoma). Of the invasive tumors included in the final analysis (n=24), all had a combined Nottingham grade of 1 or 2. Most of them were ductal carcinomas (IDC), although three tumors were classified as mucinous carcinomas. Multiple 10 µm paraffin

sections were cut from a total of 206 paraffin blocks. The intraductal and invasive tumors were microdissected using the Arcturus PixCellII and Autopix 100 laser capture microdissection (LCM) apparatuses (Applied Biosystems/Life Technologies, Carlsbad, CA). From each block, enough tumor cells were microdissected to yield at least 1 μg of genomic DNA. Tumor cellularity approached 100% for the intraductal carcinomas but was somewhat less for the invasive components (~80%). This is due to the fact that there is a limit to the resolution of the laser capture microdissection technique. Ducts involved by DCIS often have a diameter of 100 μm or more, and thus it typically is possible to excise almost pure tumor cell populations with cellularity approaching 100%. In contrast, many invasive carcinomas are composed of small clusters of tumor cells that may measure only 10–30 μm across, and thus it often is unavoidable to include a small percentage (up to 20%) of stromal cells in the microdissected invasive components.

DNA preparation

Genomic DNA was prepared from all microdissected FFPE tumor samples using the Gentra PureGene Tissue kit (Qiagen Inc., Valencia, CA) as per manufacturer's instructions. To enrich for samples with adequate DNA quality and quantity, we employed a screening assay whereby 1 to 4 μg of the tumor DNA was incubated for 3 hours using the Enzo BioScore kit (Enzo Life Sciences Inc., Farmingdale, NY). 118 of 206 DNA extracts were judged to have sufficient quantity and quality for array CGH (aCGH) analysis based on the results of the BioScore assay (threshold 1 μg). For procedural quality control, all analyses were performed as sex-mismatch hybridizations with a male control pool containing DNA from 20 cytogenetically normal individuals.

Array comparative genomic hybridization

The aCGH experiments were performed in the 2006 to 2007 time period. One μg of male control and the tumor DNA (1–5 μg of BioScore product) was random primer labeled with Cy3 and Cy5 respectively, using a BioArray CGH Labeling System (Enzo Life Sciences, Farmingdale, NY) as previously described (Desouki et al. 2010). Hybridization of the pooled probes was performed on the RPCI 19k bacterial artificial chromosome (BAC) arrays (GPL8454), followed by washing and scanning on a GenePix 4200AL Scanner (Molecular Devices Inc., Sunnyvale, CA). The arrays were comprised of ~19,000 BACs in duplicate. The 19k BAC array submission can be found as GPL8454 on the GEO web site (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8454>). Each BAC clone contained an average of 170 kb of genomic DNA. These arrays are able to detect gain or loss of one gene copy per genome and delineate the breakpoint at ~500 kb resolution. Image analysis was performed on the raw image files using the ImaGene (version 6.1.0) software (BioDiscovery Inc., El Segundo, CA). Spots of poor quality, as well as spots with signal levels indistinguishable from the background were flagged and excluded from subsequent analyses. The log₂ Cy5/Cy3 (i.e., log tumor/control) ratios for spots which passed the quality control steps were lowess corrected to account for non-linear dependency of the signal ratio to the signal intensity. Of 118 hybridization experiments, 69 yielded copy number profiles that were suitable for statistical analysis: 20 PD, 25 MD, 24 MI. Forty-nine samples yielded inadequate array data due to excessive DNA degradation and nonspecific hybridization. The evaluable samples included 18 MD-MI pairs. Seven MD samples had no evaluable invasive component, while six MI samples had no evaluable intraductal component. This group of evaluable cases did not significantly differ from unevaluable cases with regard to lesional size or time in storage.

Microarray statistical analysis

The aCGH data were normalized using the R computing language, previously described software (Gaile et al. 2007), and by the methods recently reported by us (Miecznikowski et

al. 2011). The aCGH profiles were segmented using the DNACopy R library (Olshen et al. 2004) and aberration probabilities and calls were obtained using the CGHcall R library (van de Wiel et al. 2007). Analysis was restricted to the 22 autosomal chromosomes. A principal components analysis was conducted using a feature set of 408 BACs. The feature set was created by an *ad hoc* forward selection heuristic: (step 1) segments identified by DNACopy were called aberrant using cut-off values of -0.125 and 0.1325 (see below), (step 2) the set of assayed BACs was scanned to identify the BAC which was common to the largest number of copy number aberrant segments, (step 3) the BAC identified in step 2 was added to the feature list and the segments for which it was a member were removed from further consideration by the algorithm, and (step 4) steps 2 and 3 were repeated until no copy number aberrant segments remained for consideration. Cut-offs for copy number losses and gains were determined using the following heuristics. The sample means and standard deviations of the fitted $\log_2$ T/C values were calculated across the matched and unmatched primary tumors for each BAC. A two-dimensional kernel smoothed density estimator was fitted to estimate the bivariate distribution of the sample mean and standard deviations and the sample mode was identified. Empirical cut-offs for aberrations were calculated as the sample mean corresponding to the mode plus/minus 1.96 times the sample deviation corresponding to the mode. For the purpose of these analyses, we did not further differentiate the amplitude of the copy number changes, and in this paper the terms copy number gains/amplifications and copy number losses/deletions, respectively, are used synonymously.

Genome scans comparing PD *versus* MD and MD *versus* MI aberration signatures were conducted. These scans utilized a pseudo-likelihood test statistic with a permutation derived null distribution. Permutation based P-values were calculated as well as P-values adjusted for multiple testing (Westfall et al. 1993). The pseudo-likelihood test statistic utilized a likelihood function which assumed that each BAC within each sample was a single observation on a multinomial distribution with three states (i.e., loss, normal, and gain). Since the true copy number state was unknown, the corresponding CGHcall posterior probabilities were utilized in substitution for the appropriate indicator functions. Figures were generated using software previously described by us (Gaile et al. 2010).

qPCR utilizing genomic DNA

Comparative qPCR experiments were performed to confirm the differential copy number changes of 19 candidate genes from 16 regions on 13 chromosomal arms that were differentially amplified or deleted in PD *versus* MD lesions. These regions were comprised of numerous BACs that in aggregate included over 100 different genes. The candidate genes were selected from these regions on the basis of statistical probability (low P-value) and presumed functional significance, enriching for potential tumor suppressors, nuclear antigens, and cell surface proteins. The commonly used reference gene *GAPDH* (with a very low frequency of reported copy number polymorphisms) was utilized as the normalizer and normal genomic DNA as the calibrator. The PCR protocol has previously been reported by us (Desouki et al. 2010, 2011). A 10 ng aliquot of genomic DNA (same extract as used for aCGH but not subjected to the Enzo BioScore step) was added into the PCR mixture which contained 0.1 $\mu\text{mol}/\mu\text{l}$ forward and reverse primers and 1 $\times$ Master Mix (Brilliant SYBR Green QPCR Master Mix, Stratagene, La Jolla, CA). Primer sequences for all genes are listed in Table 1. Due to the fragmented nature of the FFPE DNA samples, amplicon size was restricted to 80–100 base pairs. A reference dye (ROX) in 500-fold dilution was added to compensate for non-PCR related variations in fluorescence. Total volume for each reaction was 25 μl . The reaction was executed on a MxPro 3005P real-time PCR cycler (Stratagene). The thermal cycling conditions comprised an initial denaturation step at 95°C for 10 min followed by 40 cycles at 95°C for 30 s, 55°C for 1 min and 72°C for 30 s.

Experiments were done in duplicate for each DNA sample in a 96-well plate. The comparative fold change of the copy number for a gene was calculated by the software controlling the PCR cycler MxPro 3005P, and presented as log fold change to calibrator (log base 2). Cut-offs for amplifications and deletions were +1.0 and -0.5 ($\Delta\Delta C_T$, log2), respectively.

Development of multi-gene models of DCIS progression

Frequencies of copy number losses (L), normal (N) and gains (G) observed in patients with indolent (PD) and aggressive DCIS (MD) were reported as contingency tables, and association to DCIS progression was assessed using Fisher's exact test. A multiplex marker for DCIS status was developed from the 8 genes that differentiated PD and MD (unadjusted Fisher's exact P-value = 0.1). In order to construct a cumulative score of DNA copy number changes, the ordinal relationship between copy number and progression was assessed using the Mean Score chi-squared test, and values of -1/0/+1 were assigned to the three levels of DNA copy number according to increased frequencies in MD. A score of +1 was assigned to copy number losses for nuclear receptor corepressor 2/silencing mediator of retinoid and thyroid hormone receptor (*NCOR2/SMRT*), nuclear receptor subfamily 4 group A member 1/Nur77 orphan receptor TR3 (*NR4A1/NUR77/TR3*), dynein light chain roadblock-type 2/dynein cytoplasmic light polypeptide 2B (*DYNLRB2/DNLC2B*), suppression of tumorigenicity 13/Hsp70 interacting protein (*ST13/HIP*), uroplakin 3A (*UPK3A*) and cadherin EGF LAG seven-pass G-type receptor 1 (*CELSR1*), and for copy number gains in Grb2-related adapter protein 2 (*GRAP2*) and TATA box binding protein-associated factor 1C (*TAF1C*). The opposite copy number changes in the respective genes received a score of -1. Association of the additive score to PD/MD status was confirmed by Wilcoxon-rank sum tests.

Results

Distribution of DNA copy number aberrations in intraductal and invasive carcinomas of the breast

Pure low grade DCIS lesions without an invasive component (PD) had a higher average frequency of DNA copy number changes compared to mixed DCIS (MD) that had progressed to invasive carcinoma (Table 2, $P = 0.047$ for difference in total fractional copy number aberrations (CNAs)). This was due to a higher average frequency of gains (8.4% *versus* 6.3%, $P = 0.084$) and losses (8.2% *versus* 5.5%, $P = 0.048$). On the other hand, invasive carcinomas (MI) had similar frequencies of DNA copy number aberrations compared to the paired intraductal lesions (MD) (Table 2).

Figure 1 is a Principal Component Analysis (PCA) plot showing the genomic relationship of paired intraductal (MD) and invasive carcinomas (MI) as well as the pure DCIS lesions with no associated invasive component (PD). Shorter connecting lines between MD and MI pairs indicate a closer relationship at the DNA level. It is evident that some pairs were closely related, while others were more divergent. While the MD/MI lesions were distributed throughout the three-dimensional space, almost all of the pure DCIS lesions (PD) aggregated in a more defined area (circled area in Figure 1).

Figure 2 illustrates the distribution of DNA copy number gains and losses across the whole autosomal genome in the three groups of lesions. CNAs were identified on all autosomal arms. Among the more frequent changes were previously described common CNAs in human breast cancer including gains on 1q, 8q and 16p, as well as losses on 8p, 9p, 13q and 16q. There were statistically significant differences in copy number aberrations between pure (PD) and mixed (MD) DCIS lesions in 16 regions on 13 chromosomal arms: 1q, 3p, 5q,

6p, 9q, 10q, 11q, 12q, 16p, 16q, 18q, 20q, 22q (arrows in Figure 2). The majority of differences were represented by an increased number of copy number gains in PD lesions. Only one region showed a higher frequency of copy number changes in MD lesions, i.e., more deletions on 22q. Figure 3 shows a heatmap of this chromosomal arm. Near the telomeric end, gains were more common in the PD group. The remainder of the chromosome showed a higher frequency of deletions in the MD group.

There were no significant differences in the distribution of chromosomal copy number changes between mixed DCIS (MD) and invasive carcinomas (MI) (Figure 2).

Validation of CNAs by qPCR

Since microarray profiling experiments can be associated with substantial false positive rates, we performed qPCR studies for 19 candidate genes localized to differentially amplified or deleted chromosomal regions, utilizing aliquots of the same genomic DNA extracts that had been used for aCGH. These genes were selected based on a combination of statistical probability and potential biological interest.

Eight of these genes could be confirmed to have differential copy number changes in pure (PD) *versus* mixed DCIS (MD) (Table 3). The two genes on 12q, *SMRT* and *NR4A1*, as well as *DYNLRB2* on 16q23.2 were more commonly amplified in PD and more frequently deleted in MD. *TAF1C* (on 16q24.1) was more frequently deleted in PD. *CELSR1* and *UPK3A* (both on 22q13.31) were more commonly amplified in PD, while *GRAP2* showed more common gains in MD. *ST13* was remarkable for a higher frequency of both gains and losses in PD *versus* MD.

Eleven of the 19 candidate genes (interferon regulatory factor 8 (*IRF8*), transmembrane protein 9 (*TMEM9*), calneuron 1 (*CALN1*), centaurin alpha (*CENTA1*), F-box protein 31 (*FBXO31*), Ikaros family zinc finger 1 (*IKZF1*), fidgetin-like 1 (*FIGNL1*), THAP domain containing 7 (*THAP7*), cadherin 13 (*CDH13*), mitogen activated protein kinase 9 (*MAPK9*), ubiquitin specific peptidase 10 (*USPN10*)) could not be confirmed to have significantly different CNA frequencies in PD *versus* MD lesions. Similar to *ST13*, some of these genes showed both copy number gains and losses (15%–45%). *TMEM9*, *CALN1* and *CENTA1* were amplified in 30–40%, while *IRF8* was deleted in about 50% of all LGDCIS.

Multi-gene model of DCIS progression

For the eight confirmed candidate genes showing differential copy number changes in PD *versus* MD lesions by qPCR, we explored whether a combination of all or some of these genes may have greater discriminatory power. We used a simple binary scoring system for each gene copy number change, with a score of +1 favoring MD and a score of –1 favoring PD. For each lesion, the 8 scores were then added (theoretical range –8 to +8). A positive score indicates DCIS associated with stromal invasion (MD), while a negative score suggests DCIS without an associated invasive component (PD). The composite score from all 8 genes was highly significant in distinguishing the two types of DCIS. The median score was +1 for MD and –2 for PD lesions (Figure 4 left, Wilcoxon $P < 0.0001$). Eliminating three of the genes (*TAF1C*, *ST13*, *UPK3A*) yielded a 5-gene score that had equally high discriminatory power (median scores +1 and –1 for MD and PD, respectively; Figure 4 right, Wilcoxon $P < 0.0001$).

Discussion

Ductal carcinoma *in situ* generally is regarded as a precursor of invasive mammary carcinoma. A biopsy diagnosis of DCIS typically is followed by definitive therapy including partial or total mastectomy, often followed by adjuvant radiation and, in some cases,

systemic endocrine therapy. Due to the perceived need to eradicate DCIS diagnosed on biopsy, the natural history of this disease is poorly understood. Available evidence suggests that less than half of these lesions may progress to invasive disease, and observations from the Nashville cohort indicate that less than 30% of low grade DCIS will progress within 20 years (Sanders et al. 2005). The molecular aberrations that determine whether a given DCIS lesion has the capacity to progress have remained elusive. We investigated only low grade DCIS cases since overtreatment is a particularly significant problem in this group. In order to discover genomic aberrations associated with the propensity for stromal invasion, we employed an unbiased whole genome profiling approach. Laser capture microdissection of LGDCIS cases with and without associated stromal invasion allowed us to interrogate pure or almost pure tumor cell populations. Due to extensive degradation of DNA in FFPE breast tissues, we had to microdissect 206 paraffin blocks in order to perform 118 aCGH hybridizations, 69 of which yielded data that were amenable to statistical analysis. Compared to newer platforms some of which were not available when our experiments were performed, BAC arrays were found to be more robust, have better signal-to-noise ratios, and provide a more accurate measure of copy number changes in poor quality DNA (Nowak et al. 2007). Our DNA profiling experiments revealed many DNA copy number changes that were previously described in low grade mammary neoplasia, including gains on 1q and 8q, as well as losses on 8p, 9p and 16q (Fujii et al. 1996; James et al. 1997). This confirms the ability of our BAC arrays to detect expected CNAs and indicates good sensitivity.

We identified no significant differences in the quantity or quality of DNA copy number aberrations in invasive carcinomas compared with paired intraductal lesions (Table 2, Figure 2). This is consistent with a substantial body of work that has failed to reveal non-random differences between *in situ* and paired invasive breast cancers at the DNA, RNA and protein level (O'Connell et al. 1998; Buerger et al. 1999; Aubele et al. 2000; Leong et al. 2001; Farabegoli et al. 2002; Burstein et al. 2004; Hwang et al. 2004; Simpson et al. 2005; Castro et al. 2008; Iakovlev et al. 2008). Our observation supports the notion that the biological potential of a mammary neoplasm is determined at the *in situ* carcinoma stage (or earlier) (Burstein et al. 2004).

A major finding was the identification of specific DNA copy number gains and losses that distinguished LGDCIS lesions with and without an associated invasive component. Pure DCIS had a larger number of CNAs than DCIS associated with invasion (Table 2, Figure 2). While this observation may seem counterintuitive, it is not unprecedented. Farabegoli *et al.* described a higher proportion of molecular abnormalities in non-progressing *versus* progressing DCIS lesions (Farabegoli et al. 2002). Employing an aCGH approach similar to ours, albeit on a smaller number and a more heterogeneous group of samples, Iakovlev *et al.* also reported higher genomic complexity in non-progressing DCIS lesions (Iakovlev et al. 2008). Using a gene expression profiling approach on a limited number of morphologically heterogeneous (including many high grade) lesions, Castro *et al.* described 147 genes that were differentially expressed in PD *versus* MD lesions, with the majority of the genes showing upregulation in pure DCIS (Castro et al. 2008). In contrast, O'Connell *et al.* reported a higher LOH frequency at 2p, 11p and 17q in DCIS associated with invasive tumors compared to pure DCIS (O'Connell et al. 1998). However, the latter study used PCR-based methodology to interrogate a wide variety of DCIS lesions including many high grade neoplasms, which are known to be molecularly distinct from low grade lesions. Interestingly, we found that many of the chromosomal aberrations more commonly seen in PD lesions were copy number gains, suggesting the novel concept that, at the molecular level, LGDCIS without the propensity to progress are dead end lesions that may be characterized by amplification and possible activation of invasion suppressor genes. In this scenario of active suppression, "the brakes are on". Conversely, the same or different invasion suppressor genes may be inactivated in aggressive DCIS lesions ("brakes off"). At

least 3 of the candidate suppressor genes we investigated (*NCOR2/SMRT*, *NR4A1*, *DYNLRB2*, see below) have the potential to serve as molecular switches that determine the invasive potential of a LGDCIS lesion. Principal component analysis provided additional evidence that pure DCIS lesions were qualitatively distinct from morphologically similar lesions that had progressed to invasive carcinoma (MD) (Figure 1). This observation is consistent with reports by other investigators on differential molecular aberrations in DCIS lesions with and without the capacity to progress (Castro et al. 2008; Iakovlev et al. 2008). Unlike prior studies utilizing conventional CGH, our high resolution BAC arrays afforded us the opportunity to identify candidate genes residing at the differentially amplified or deleted loci.

One of the limitations of microarray technology is a relatively high false positive rate (Pawitan et al. 2005). Thus, apparent alterations require confirmation by at least one alternate technique. Such confirmatory studies frequently are absent from publications on DNA copy number changes in human neoplasms. The 16 chromosomal regions with differential copy number changes contained a multitude of candidate genes that potentially could be involved in the promotion or suppression of stromal invasion. We selected 19 candidate genes for confirmatory studies employing comparative qPCR. These candidate genes were selected based on statistical probability and potential biological function. Of the 19 genes tested, 8 were found to be differentially amplified or deleted in PD *versus* MD lesions. This confirmation rate (42%) is within the range of what was previously reported by other investigators (Castro et al. 2008; Wang et al. 2009). It actually is better than the confirmation rate (30%) in our previously published study on more advanced breast cancers (Desouki 2010, 2011). The latter study utilized DNA from macrodissected tissues yielding lower tumor cellularity and did not employ the Enzo BioScore screening step, which may explain the higher failure rate. In general, many investigators have found DNA extracted from FFPE breast tissues to be an extremely challenging analyte.

Table 4 provides a summary of what currently is known about the function of the 8 candidate genes and their involvement in human tumors. Five of the genes were more commonly amplified in pure DCIS, and three of these additionally were more commonly deleted in DCIS lesions that had progressed to invasive disease (MD), suggesting that they may function as invasion suppressor genes. While the flamingo-type cadherin *CELSR1* (on 22q13.31) has no known role in human neoplasia, the other four genes reportedly are downregulated in a variety of human tumors. *SMRT/NCOR2* (on 12q31) previously was found to be downregulated in multiple myeloma (Ghoshal et al. 2009), and loss of expression in non-Hodgkin's lymphoma was associated with transformation and progression to a more aggressive phenotype (Song et al. 2005). Its role in human breast cancer is controversial. The gene is thought to be required for Tamoxifen to exert its full anti-estrogenic action in breast tumor cells (Fleming et al. 2004; Peterson et al. 2007). While Green *et al.* described that SMRT-positive breast cancers followed a more aggressive clinical course (Green et al. 2008), the opposite was reported by van Agthoven *et al.* (van Agthoven et al. 2009). *NR4A1/NUR77/TR3* (on 12q13.13) previously was shown to be downregulated in acute myelogenous leukemia, metastatic melanoma and follicular thyroid carcinoma (Mullican et al. 2007; Yu et al. 2007; Camacho et al. 2009). Alexopoulou *et al.* described lower expression levels of this gene in high grade and metastatic breast tumors (Alexopoulou et al. 2010). Our finding of differential copy number changes in these two genes in LGDCIS lesions suggests that they may play a role not only in the progression of established breast cancers, but additionally they may determine whether a mammary neoplasm invades the surrounding stroma in the first place. *DYNLRB2/DNLC2B* (on 16q23.2) encodes a component of the dynein light chain and thus may be involved in cellular migration (Wanschers et al. 2008). It was shown to be downregulated in about two thirds of hepatocellular carcinomas (Jiang et al. 2001). *UPK3A* (on 22q13.31) encodes a cell

membrane protein involved in urothelial differentiation, and downregulation of its expression was associated with worse prognosis in transitional cell carcinomas (Huang et al. 2007; Matsumoto et al. 2008). Our data newly implicate these two genes in early human mammary carcinogenesis.

Two of the eight candidate genes may have invasion promoting activity. The transcription co-factor *TAF1C* (on 16q24.1) was more commonly deleted in pure DCIS lesions. This gene has no known role in human neoplasia. *GRAP2* (on 22q13.1) was more commonly amplified in progressive DCIS lesions (MD). This gene encodes a RET-binding protein involved in T-cell receptor signaling (Ludwig et al. 2003). It was reported to be expressed in medullary thyroid carcinoma (Ludwig et al. 2009). Finally, *ST13* (on 22q13.1) was unique in that it showed both more common gains and more common losses in PD lesions. This finding may be somewhat analogous to the well documented presence of both amplifications and deletions in the topoisomerase II α gene in breast tumors (Di Leo et al. 2011). *ST13* is an Hsp70 co-chaperone that previously was shown to be downregulated in colon cancers (Wang et al. 2005; Shi et al. 2007). In sum, previously published observations are consistent with the notion that our eight validated genes may be functionally important in determining whether or not a LGDCIS lesion progresses to invasive carcinoma. However, it is acknowledged that the functional significance of any of these genes in early breast cancer progression has not been demonstrated.

While the eight genes described above individually had differential CNAs that distinguished between PD and MD lesions by qPCR (Table 3), we reasoned that a combination of some or all candidate genes would have greater predictive power. We developed a simple statistical model in which the most common CNA in a given gene was assigned a value of +1 if it favored MD, and -1 if it favored PD. Thus, an aggregate positive score would indicate a potentially aggressive lesion (MD), while a negative combined score would indicate indolent disease (PD). Figure 4 illustrates that the full 8-gene score indeed discriminated between PD and MD lesions with high accuracy. Interestingly, the 5-gene score excluding *TAF1C*, *ST13* and *UPK3A* had equally high discriminatory power. These results demonstrate the potential predictive value of a combined score, but a complete evaluation of performance requires validation in independent samples. We currently are pursuing the possibility that these findings may provide the basis of a generally applicable tissue-based multi-gene assay that differentiates indolent DCIS lesions from those that require definitive therapy.

In conclusion, our study provides compelling evidence that low grade DCIS lesions with and without associated invasion have distinct DNA profiles that may be exploited as a prognostic tool. Pure DCIS lesions contain a larger number of CNAs, and many of them may serve to actively suppress stromal invasion. We identified 8 candidate genes with differential copy number changes in indolent (PD) *versus* aggressive DCIS (MD) lesions. It remains unknown if these genes drive stromal invasion or whether they are bystanders. Either way, they may prove to be useful biomarkers, and their differential patterns of amplification or deletion may provide the basis for new assays identifying DCIS lesions that do not require aggressive treatment.

Acknowledgments

This work was supported by grants from the NIH (R21CA106676) and Friends for an Earlier Breast Cancer Test (both to JG). The Genomics Shared Resource at Roswell Park Cancer Institute is a Cancer Center Support Grant (CCSG) Shared Resource (NIH 5P30CA016056-34). We thank Devin McQuaid for expert technical assistance.

References

- Alexopoulou AN, Leao M, Caballero OL, Da Silva L, Reid L, Lakhani SR, Simpson AJ, Marshall JF, Neville AM, Jat PS. Dissecting the transcriptional networks underlying breast cancer: NR4A1 reduces the migration of normal and breast cancer cell lines. *Breast Cancer Res.* 2010; 12:R51. [PubMed: 20642837]
- Aubele M, Mattis A, Zitzelsberger H, Walch A, Kremer M, Welzl G, Hofler H, Werner M. Extensive ductal carcinoma In situ with small foci of invasive ductal carcinoma: evidence of genetic resemblance by CGH. *Int J Cancer.* 2000; 85:82–86. [PubMed: 10585588]
- Bombonati A, Sgroi DC. The molecular pathology of breast cancer progression. *J Pathol.* 2011; 223:307–317. [PubMed: 21125683]
- Buerger H, Otterbach F, Simon R, Poremba C, Diallo R, Decker T, Riethdorf L, Brinkschmidt C, Dockhorn-Dworniczak B, Boecker W. Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. *J Pathol.* 1999; 187:396–402. [PubMed: 10398097]
- Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. *N Engl J Med.* 2004; 350:1430–1441. [PubMed: 15070793]
- Camacho CP, Latini FR, Oler G, Hojaij FC, Maciel RM, Riggins GJ, Cerutti JM. Down-regulation of NR4A1 in follicular thyroid carcinomas is restored following lithium treatment. *Clin Endocrinol.* 2009; 70:475–483.
- Castro NP, Osorio CA, Torres C, Bastos EP, Mourao-Neto M, Soares FA, Brentani HP, Carraro DM. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. *Breast Cancer Res.* 2008; 10:R87. [PubMed: 18928525]
- Desouki MM, Liao S, Huang H, Conroy J, Nowak NJ, Shepherd L, Gaile DP, Geradts J. Identification of metastasis-associated breast cancer genes using a high-resolution whole genome profiling approach. *J Cancer Res Clin Oncol.* 2010; 137:795–809. [PubMed: 20680643]
- Desouki MM, Liao S, Conroy J, Nowak NJ, Shepherd L, Gaile DP, Geradts J. The genomic relationship between primary breast carcinomas and their nodal metastases. *Cancer Invest.* 2011; 29:300–307. [PubMed: 21469979]
- Di Leo A, Desmedt C, Bartlett JM, Piette F, Ejlersen B, Pritchard KI, Larsimont D, Poole C, Isola J, Earl H, Mouridsen H, O'Malley FP, Cardoso F, Tanner M, Munro A, Twelves CJ, Sotiriou C, Shepherd L, Cameron D, Piccart MJ, Buyse M. HER2 and TOP2A as predictive markers for anthracycline-containing chemotherapy regimens as adjuvant treatment of breast cancer: a meta-analysis of individual patient data. *Lancet Oncol.* 2011; 12:1134–1142. [PubMed: 21917518]
- Di Pietro C, Rapisarda A, Amico V, Bonaiuto C, Viola A, Scalia M, Motta S, Amato A, Engel H, Messina A, Sichel G, Grzeschik K, Purrello M. Genomic localization of the human genes TAF1A, TAF1B and TAF1C, encoding TAF(I)48, TAF(I)63 and TAF(I)110 subunits of class I general transcription initiation factor SL1. *Cytogenet Cell Genet.* 2000; 89:133–136. [PubMed: 10894955]
- Eusebi V, Feudale E, Foschini MP, Micheli A, Conti A, Riva C, Di Palma S, Rilke F. Long-term follow-up of in situ carcinoma of the breast. *Semin Diagn Pathol.* 1994; 11:223–235. [PubMed: 7831534]
- Farabegoli F, Champeme MH, Bieche I, Santini D, Ceccarelli C, Derenzini M, Lidereau R. Genetic pathways in the evolution of breast ductal carcinoma in situ. *J Pathol.* 2002; 196:280–286. [PubMed: 11857490]
- Fleming FJ, Hill AD, McDermott EW, O'Higgins NJ, Young LS. Differential recruitment of coregulator proteins steroid receptor coactivator-1 and silencing mediator for retinoid and thyroid receptors to the estrogen receptor-estrogen response element by beta-estradiol and 4-hydroxytamoxifen in human breast cancer. *J Clin Endocrinol Metab.* 2004; 89:375–383. [PubMed: 14715875]
- Frykberg ER, Bland KI. Overview of the biology and management of ductal carcinoma in situ of the breast. *Cancer.* 1994; 74(1 Suppl):350–361. [PubMed: 8004607]
- Fujii H, Szumel R, Marsh C, Zhou W, Gabrielson E. Genetic progression, histological grade, and allelic loss in ductal carcinoma in situ of the breast. *Cancer Res.* 1996; 56:5260–5265. [PubMed: 8912866]

- Gaile, DPMJ.; Shepherd, LS. aCGHplus: an R library for the storage, analysis and visualization of aCGH data. Source code and documentation. 2007. available at: <http://sphhp.buffalo.edu/biostat/research/software/acghplus/index.php>
- Gaile DPSL, Bruno AE, Liu S, Morrison C, Sucheston L, Miecznikowski J. iGenomicViewer: R package for visualisation of high dimension genomic data. *Int J Bioinform Res Appl*. 2010; 6:10.
- Ghoshal P, Nganga AJ, Moran-Giusti J, Szafrank A, Johnson TR, Bigelow AJ, Houde CM, Avet-Loiseau H, Smiraglia DJ, Ersing N, Chanan-Khan AA, Coignet LJ. Loss of the SMRT/NCOR2 corepressor correlates with JAG2 overexpression in multiple myeloma. *Cancer Res*. 2009; 69:4380–4387. [PubMed: 19417136]
- Green AR, Burney C, Granger CJ, Paish EC, El-Sheikh S, Rakha EA, Powe DG, Macmillan RD, Ellis IO, Stylianou E. The prognostic significance of steroid receptor co-regulators in breast cancer: co-repressor NCOR2/SMRT is an independent indicator of poor outcome. *Breast Cancer Res Treat*. 2008; 110:427–437. [PubMed: 17902051]
- Hadjantonakis AK, Sheward WJ, Harmar AJ, de Galan L, Hoovers JM, Little PF. Celsr1, a neural-specific gene encoding an unusual seven-pass transmembrane receptor, maps to mouse chromosome 15 and human chromosome 22qter. *Genomics*. 1997; 45:97–104. [PubMed: 9339365]
- Huang HY, Shariat SF, Sun TT, Lepor H, Shapiro E, Hsieh JT, Ashfaq R, Lotan Y, Wu XR. Persistent uroplakin expression in advanced urothelial carcinomas: implications in urothelial tumor progression and clinical outcome. *Hum Pathol*. 2007; 38:1703–1713. [PubMed: 17707461]
- Hwang ES, DeVries S, Chew KL, Moore DH 2nd, Kerlikowske K, Thor A, Ljung BM, Waldman FM. Patterns of chromosomal alterations in breast ductal carcinoma in situ. *Clin Cancer Res*. 2004; 10:5160–5167. [PubMed: 15297420]
- Iakovlev VV, Arneson NC, Wong V, Wang C, Leung S, Iakovleva G, Warren K, Pintilie M, Done SJ. 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; 14:4446–4454. [PubMed: 18628458]
- James LA, Mitchell EL, Menasce L, Varley JM. Comparative genomic hybridisation of ductal carcinoma in situ of the breast: identification of regions of DNA amplification and deletion in common with invasive breast carcinoma. *Oncogene*. 1997; 14:1059–1065. [PubMed: 9070654]
- Jiang J, Yu L, Huang X, Chen X, Li D, Zhang Y, Tang L, Zhao S. Identification of two novel human dynein light chain genes, DNLC2A and DNLC2B, and their expression changes in hepatocellular carcinoma tissues from 68 Chinese patients. *Gene*. 2001; 281:103–113. [PubMed: 11750132]
- Leong AS, Sormunen RT, Vinyuvat S, Hamdani RW, Suthipintawong C. Biologic markers in ductal carcinoma in situ and concurrent infiltrating carcinoma. A comparison of eight contemporary grading systems. *Am J Clin Pathol*. 2001; 115:709–718. [PubMed: 11345835]
- Li QX, Ke N, Sundaram R, Wong-Staal F. NR4A1, 2, 3—an orphan nuclear hormone receptor family involved in cell apoptosis and carcinogenesis. *Histol Histopathol*. 2006; 21:533–540. [PubMed: 16493583]
- Ludwig L, Kessler H, Hoang-Vu C, Dralle H, Adler G, Boehm BO, Schmid RM. Grap-2, a novel RET binding protein, is involved in RET mitogenic signaling. *Oncogene*. 2003; 22:5362–5366. [PubMed: 12917638]
- Ludwig L, Oswald F, Hoang-Vu C, Dralle H, Hildt E, Schmid RM, Karges W. Expression of the Grb2-related RET adapter protein Grap-2 in human medullary thyroid carcinoma. *Cancer Lett*. 2009; 275:194–197. [PubMed: 19027225]
- Matsumoto K, Satoh T, Irie A, Ishii J, Kuwano S, Iwamura M, Baba S. Loss expression of uroplakin III is associated with clinicopathologic features of aggressive bladder cancer. *Urology*. 2008; 72:444–449. [PubMed: 18313120]
- Miecznikowski JCGD, Liu S, Shepherd L, Nowak N. A new normalizing algorithm for BAC CGH arrays with quality control metrics. *J Biomed Biotechnol*. 2011 Epub ahead of print Feb. 21.
- Mullican SE, Zhang S, Konopleva M, Ruvolo V, Andreeff M, Milbrandt J, Conneely OM. Abrogation of nuclear receptors Nr4a3 and Nr4a1 leads to development of acute myeloid leukemia. *Nat Med*. 2007; 13:730–735. [PubMed: 17515897]

- Nowak NJ, Miecznikowski J, Moore SR, Gaile D, Bobadilla D, Smith DD, Kernstine K, Forman SJ, Mhawech-Fauceglia P, Reid M, Stoler D, Loree T, Rigual N, Sullivan M, Weiss LM, Hicks D, Slovak ML. Challenges in array comparative genomic hybridization for the analysis of cancer samples. *Genet Med*. 2007; 9:585–595. [PubMed: 17873646]
- O'Connell P, Pekkel V, Fuqua SA, Osborne CK, Clark GM, Allred DC. Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J Natl Cancer Inst*. 1998; 90:697–703. [PubMed: 9586667]
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004; 5:557–572. [PubMed: 15475419]
- Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics*. 2005; 21:3017–3024. [PubMed: 15840707]
- Peterson TJ, Karmakar S, Pace MC, Gao T, Smith CL. The silencing mediator of retinoic acid and thyroid hormone receptor (SMRT) corepressor is required for full estrogen receptor alpha transcriptional activity. *Mol Cell Biol*. 2007; 27:5933–5948. [PubMed: 17591692]
- Polyak K, Hu M. Do myoepithelial cells hold the key for breast tumor progression? *J Mammary Gland Biol Neoplasia*. 2005; 10:231–247. [PubMed: 16807803]
- Sanders ME, Schuyler PA, Dupont WD, Page DL. 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*. 2005; 103:2481–2484. [PubMed: 15884091]
- Shi ZZ, Zhang JW, Zheng S. What we know about ST13, a co-factor of heat shock protein, or a tumor suppressor? *J Zhejiang Univ Sci B*. 2007; 8:170–176. [PubMed: 17323428]
- Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ, Poller DN. A prognostic index for ductal carcinoma in situ of the breast. *Cancer*. 1996; 77:2267–2274. [PubMed: 8635094]
- Simpson PT, Reis-Filho JS, Gale T, Lakhani SR. Molecular evolution of breast cancer. *J Pathol*. 2005; 205:248–254. [PubMed: 15641021]
- Song L, Zlobin A, Ghoshal P, Zhang Q, Houde C, Weijzen S, Jiang Q, Nacheva E, Yagan D, Davis E, Galiegue-Zouitina S, Catovsky D, Grogan T, Fisher RI, Miele L, Coignet LJ. Alteration of SMRT tumor suppressor function in transformed non-Hodgkin lymphomas. *Cancer Res*. 2005; 65:4554–4561. [PubMed: 15930272]
- van Agthoven T, Sieuwerts AM, Veldscholte J, Meijer-van Gelder ME, Smid M, Brinkman A, den Dekker AT, Leroy IM, van Ijcken WF, Sleijfer S, Foekens JA, Dorssers LC. CITED2 and NCOR2 in anti-oestrogen resistance and progression of breast cancer. *Br J Cancer*. 2009; 101:1824–1832. [PubMed: 19904269]
- van de Wiel MAKK, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics*. 2007; 23:3.
- Wang C, Iakovlev VV, Wong V, Leung S, Warren K, Iakovleva G, Arneson NC, Pintilie M, Miller N, Youngson B, McCready DR, Done SJ. Genomic alterations in primary breast cancers compared with their sentinel and more distal lymph node metastases: an aCGH study. *Genes Chromosomes Cancer*. 2009; 48:1091–1101. [PubMed: 19760610]
- Wang LB, Zheng S, Zhang SZ, Peng JP, Ye F, Fang SC, Wu JM. Expression of ST13 in colorectal cancer and adjacent normal tissues. *World J Gastroenterol*. 2005; 11:336–339. [PubMed: 15637739]
- Wanschers B, van de Vorstenbosch R, Wijers M, Wieringa B, King SM, Fransen J. Rab6 family proteins interact with the dynein light chain protein DYNLRB1. *Cell Motil Cytoskeleton*. 2008; 65:183–196. [PubMed: 18044744]
- Westfall, PHYS. Resampling-based multiple testing: examples and methods for p-value adjustment. Wiley; New York: 1993.
- Yu H, Kumar SM, Fang D, Acs G, Xu X. Nuclear orphan receptor TR3/Nur77 mediates melanoma cell apoptosis. *Cancer Biol Ther*. 2007; 6:405–412. [PubMed: 17297306]
- Yu J, Palmer C, Alenghat T, Li Y, Kao G, Lazar MA. The corepressor silencing mediator for retinoid and thyroid hormone receptor facilitates cellular recovery from DNA double-strand breaks. *Cancer Res*. 2006; 66:9316–9322. [PubMed: 16982777]

Figure 1. 3-D principal components plot for 20 PD (gray squares), 25 MD (open circles), and 24 MI (closed circles) lesions

Plotted coordinates for each sample correspond to the loadings for the three largest principal components. 18 cases had both an MD and an MI component (paired lesions connected by black lines). Most of the PD cases clustered separately (oval).

Figure 2. DNA copy number changes in PD (n=20) versus MD (n=25) and MI (n=24) lesions
 The autosomal DNA copy number gains (positive, black) and losses (negative, gray) in all microdissected lesions are plotted across the genome from 1p (left) to 22q (right). Black arrows indicate regions of copy number changes that were statistically more common in the PD group. The dotted arrow indicates the sole region that was more commonly deleted in the MD group. In contrast, there were no significant copy number changes that differentiated mixed DCIS lesions (MD) from their matched invasive components (MI).

Figure 3. Heat map of chromosomal arm 22q

The x-axis shows the 20 PD lesions (pink) on the left and the 25 MD lesions (purple) on the right. The scale on the far right indicates the $-\log_{10}$ P-values for differential amplification (red) or deletion (green) in the two groups. The y-axis shows the chromosomal bands from telomere (top) to centromere (bottom). Copy number gains are in red, losses are in green, and normal copy numbers are in black. Near the telomeric end, gains were more common in the PD group. The remainder of the chromosome showed a higher frequency of deletions in the MD group.

Figure 4. Aggregate gene score as discriminator of aggressive (MD, n=25) and indolent (PD, n=20) LGDCIS cases

Numerical values (+1, 0, -1) were assigned to copy numbers in each gene, with positive numbers favoring MD and negative numbers favoring PD. Box plots are shown for the full 8-gene score (left) and the 5-gene score (right). The 5-gene score excludes *TAF1C*, *ST13* and *UPK3A*. For both scores, the P-value (Mean Score chi-squared test) for the difference between PD and MD was <0.0001.

Table 1

Quantitative PCR primer data for genes included in BACs with differential copy number changes in pure DCIS (PD) *versus* mixed DCIS samples (MD)

Genes	Primer Designation	Nucleotide Sequence	Amplicon (base pairs)
Candidate Genes			
<i>SMRT (NCOR2)*</i>	Forward	TCCCAGCGTGACTCTTTTCT	
	Reverse	ACCGGTGCTATGAACCAGTC	91
<i>NR4A1 (NUR77, TR3)*</i>	Forward	CTGAGACCACCAGGAAGAGC	
	Reverse	TCCTCCTCACGATCTGCTTT	99
<i>DYNLRB2 (DNLC2B)*</i>	Forward	GGGTAGCGTTGTTGACATCC	
	Reverse	ACGGAGACCCAGATTTACC	90
<i>TAF1C*</i>	Forward	GCAGGTCCCTCAGTTCTCAC	
	Reverse	ACTCCACCCTACCAAACAC	84
<i>CELSR1*</i>	Forward	GGTGTCTAGCAGTGGGTGT	
	Reverse	AAAGCTGGAGGCAAAAGACA	90
<i>UPK3A*</i>	Forward	GCCCACATTTTCATTTTGCT	
	Reverse	TGTAAGACGAAGGGGTTTGG	100
<i>ST13(HIP)*</i>	Forward	CCCCGACAGCTGTAATTCAT	
	Reverse	GTCGAGCACAACTTCAGCA	88
<i>GRAP2*</i>	Forward	CCTGTGTGGCATTACCCTCT	
	Reverse	CCCCTTTAAGAGCTGTCACG	80
<i>CALN1</i>	Forward	GAAAAGCGGCTCTGTCAAAC	
	Reverse	GCGCTCTGTGAATTTGTGAA	96
<i>CDH13</i>	Forward	ATTCACAGGAGTGGGTTTGC	
	Reverse	ATTTCTTGCCAAGGTCATGC	89
<i>CENTA1</i>	Forward	CCAGCCGCAATTTTGTATT	
	Reverse	TGGAGGATCAGGAGTTGGAC	80
<i>FBXO31</i>	Forward	GGGTCGTATTCACGCTGTTT	
	Reverse	CGCGAGGTTTATACCTGAA	90
<i>FIGNL1</i>	Forward	AGTGGTGGGAGCAACAAATC	
	Reverse	AAGCTTCTGGGAGGGGAATA	92
<i>THAP7</i>	Forward	GGACAACCAAGACCAAAGGA	
	Reverse	CCCCACCCATATCACATACC	95
<i>IKZF1</i>	Forward	GTGTGGACCCTTCTTTTGGA	
	Reverse	GGAGCTGTGCAGCAATATGA	88
<i>IRF8</i>	Forward	ATGGAGGGTTTGTCTGTGC	
	Reverse	GAGGGTGGGAACAGCATCTA	86
<i>MAPK9</i>	Forward	CCATGTTGGATTCTTGCTT	
	Reverse	TCTCCCCTGTGGCAATTTAG	100
<i>TMEM9</i>	Forward	GGGCCTCAGCTAAGACACAG	
	Reverse	CACCAAAGGCCAGTGATTTT	88

Genes	Primer Designation	Nucleotide Sequence	Amplicon (base pairs)
Reference gene	Forward	GGCCTCCCATTCTTTTCTC	82
	Reverse	CCCAACATTCTGTGACGTG	
	Forward	CACGTATTCCCCCAGGTTTA	90
	Reverse	GACAAGCTTCCCGTTCTCAG	

* The eight genes comprising the “invasion signature panel”.

Table 2

DNA copy number aberrations (CNAs) in low grade DCIS with and without associated invasive carcinoma

	N	Average fractional gain $\pm$ SD	Average fractional loss $\pm$ SD	Average total fractional CNAs $\pm$ SD
DCIS without inv. ca. [PD]	20	0.084 $\pm$ 0.035	0.082 $\pm$ 0.050	0.166 $\pm$ 0.081
DCIS with inv. ca. [MD]	25	0.063 $\pm$ 0.041	0.055 $\pm$ 0.038	0.119 $\pm$ 0.074
Matched inv. ca. [MI]	24	0.069 $\pm$ 0.043	0.061 $\pm$ 0.034	0.130 $\pm$ 0.068
<i>TOTAL</i>	<i>69</i>	<i>0.071$\pm$0.040</i>	<i>0.065$\pm$0.042</i>	<i>0.137$\pm$0.075</i>

SD = standard deviation

Table 3

Quantitative PCR copy number changes in 8 candidate genes

	<i>SMRT (NCOR2)</i>		<i>NR4A1 (NUR77, TR3)</i>		<i>DYNLRB2 (DNLC2B)</i>		<i>GRAP2</i>		<i>CELSRI</i>		<i>UPK3A</i>		<i>TAF1C</i>		<i>ST13 (HIP)</i>	
	PD	MD	PD	MD	PD	MD	PD	MD	PD	MD	PD	MD	PD	MD	PD	MD
Loss	6 (30%)	22 (88%)	6 (30%)	18 (72%)	1 (5%)	6 (24%)	2 (11%)	0 (0%)	0 (0%)	0 (0%)	1 (5%)	2 (8%)	7 (39%)	2 (8%)	4 (20%)	1 (4%)
Normal	10 (50%)	3 (12%)	8 (40%)	7 (28%)	4 (20%)	12 (48%)	5 (28%)	2 (8%)	0 (0%)	5 (20%)	4 (20%)	12 (48%)	7 (39%)	18 (72%)	8 (40%)	21 (84%)
Gain	4 (20%)	0 (0%)	6 (30%)	0 (0%)	15 (75%)	7 (28%)	11 (61%)	23 (92%)	20 (100%)	20 (80%)	15 (75%)	11 (44%)	4 (22%)	5 (20%)	8 (40%)	3 (12%)
P-value	0.0002		0.0014		0.0110		0.0402		0.0562		0.1053		0.0308		0.0056	

Differential copy number changes in low grade DCIS with (MD) and without (PD) associated stromal invasion: Eight candidate genes validated by qPCR

Table 4

Candidate Gene	Chr. Loc.	Differential CNA (qPCR)	Function	Involvement in Human Neoplasia	References
<i>SMRT (NCOR2)</i>	12q24.31	More commonly amplified in PD, more commonly deleted in MD	Component of repressing nuclear receptor complex; promotes chromatin condensation; involved in DNA repair	Downregulated in multiple myeloma; downregulation associated with transformation of non-Hodgkin's lymphoma; expression correlated with breast cancer recurrence	Fleming et al. 2004; Song et al. 2005; Yu et al. 2006; Peterson et al. 2007; Green et al. 2008; Ghoshal et al. 2009; van Agthoven et al. 2009
<i>NR4A1 (Nur77, TR3)</i>	12q13.13	More commonly amplified in PD, more commonly deleted in MD	Nuclear orphan hormone receptor; pro-apoptotic in mitochondria, anti-apoptotic in nucleus	Downregulated in acute myelogenous leukemia, follicular thyroid carcinoma, and metastatic melanoma; expression level inversely correlated with grade and metastasis in breast cancer	Li et al. 2006; Mullican et al. 2007; Yu et al. 2007; Camacho et al. 2009; Alexopoulos et al. 2010
<i>DYNLRB2 (DNL C2B)</i>	16q23.2	More commonly amplified in PD, more commonly deleted in MD	Dynein light chain protein	Downregulated in hepatocellular carcinoma	Jiang et al. 2001; Wanschers et al. 2008
<i>TAF1C</i>	16q24.1	More commonly deleted in PD	Transcription co-factor	Not described	Di Pietro et al. 2000
<i>CELSR1</i>	22q13.31	More commonly amplified in PD	Flamingo-type cadherin; component of Wnt signaling pathway	Not described	Hadjantonakis et al. 1997
<i>UPK3A</i>	22q13.31	More commonly amplified in PD	Cell membrane protein; involved in urothelial differentiation	Downregulation associated with worse prognosis in bladder cancer	Huang et al. 2007; Matsumoto et al. 2008
<i>ST13 (HIP)</i>	22q13.1	More commonly amplified or deleted in PD	Hsp70 co-chaperone	Downregulated in colon cancer	Wang et al. 2005; Shi et al. 2007
<i>GRAP2</i>	22q13.1	More commonly amplified in MD	RET-binding protein; involved in T-cell receptor signaling	Expressed in medullary thyroid carcinoma	Ludwig et al. 2003; Ludwig et al. 2009