Research article

Characterization of adjacent breast tumors using oligonucleotide microarrays

Meredith A Unger*†, Mazhar Rishi‡, Virginia B Clemmer§, Jennifer L Hartman¶, Elizabeth A Keiper¶, Joel D Greshock*†, Lewis A Chodosh¶**, Michael N Liebman† and Barbara L Weber*†

*University of Pennsylvania Cancer Center, Philadelphia, Pennsylvania, USA
†Abramson Family Cancer Research Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
‡Department of Pathology, St Francis Hospital, Wilmington, Delaware, USA
§Department of Surgery, St Francis Hospital, Wilmington, Delaware, USA
¶Department of Molecular and Cellular Engineering, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
**Division of Endocrinology, Diabetes and Metabolism, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

Correspondence: Barbara L Weber, University of Pennsylvania Cancer Center, 514 BRB 2/3, 421 Curie Blvd, Philadelphia, PA 19104, USA. Tel: +1 215 898 0248; fax: +1 215 573 2486; e-mail: weberb@mail.med.upenn.edu

Received: 6 February 2001
Revisions requested: 9 March 2001
Revisions received: 6 April 2001
Accepted: 17 May 2001
Published: 13 July 2001

Breast Cancer Res 2001, **3**:336–341

(Print ISSN 1465-5411; Online ISSN 1465-542X)

Abstract

Background: Current methodology often cannot distinguish second primary breast cancers from multifocal disease, a potentially important distinction for clinical management. In the present study we evaluated the use of oligonucleotide-based microarray analysis in determining the clonality of tumors by comparing gene expression profiles.

Method: Total RNA was extracted from two tumors with no apparent physical connection that were located in the right breast of an 87-year-old woman diagnosed with invasive ductal carcinoma (IDC). The RNA was hybridized to the Affymetrix Human Genome U95A Gene Chip® (12,500 known human genes) and analyzed using the Gene Chip Analysis Suite® 3.3 (Affymetrix, Inc, Santa Clara, CA, USA) and JMPIN® 3.2.6 (SAS Institute, Inc, Cary, NC, USA). Gene expression profiles of tumors from five additional patients were compared in order to evaluate the heterogeneity in gene expression between tumors with similar clinical characteristics.

Results: The adjacent breast tumors had a pairwise correlation coefficient of 0.987, and were essentially indistinguishable by microarray analysis. Analysis of gene expression profiles from different individuals, however, generated a pairwise correlation coefficient of 0.710.

Conclusion: Transcriptional profiling may be a useful diagnostic tool for determining tumor clonality and heterogeneity, and may ultimately impact on therapeutic decision making.

Keywords: breast cancer, genomics, microarrays, transcriptional profiling, tumor clonality

Introduction

Appropriate clinical treatment of an individual with multiple adjacent breast tumors can be difficult to determine. If the tumors originated from the same clone, then they are likely to have common genetic changes and behave in a similar manner. Alternatively, if the tumors are separate primary lesions, then different genetic pathways might have been disrupted and the tumors may require separate treatment strategies [1]. Current methods to detect tumor clonality include methylation-sensitive restriction enzyme digestions and detection of X chromosome polymorphisms, neither of which give an overall genomic assessment of tumors [2].

ER = estrogen receptor; IDC = invasive ductal carcinoma; PR = progesterone receptor; VEGF = vascular endothelial growth factor.

We tested the capabilities of oligonucleotide microarrays in determining tumor clonality. We obtained samples from adjacent tumors from a single individual, and compared the transcriptional profile of each tumor. For comparison, we determined the gene expression profiles of tumors with similar clinical characteristics from five different individuals. These findings were interpreted in the context of the observed chip-to-chip variation and with duplicate labeling of the same RNA source.

Patients and method

Study subjects and selection criteria

Two foci of IDC were removed from the right breast of an 87-year-old woman (tumors 5A and 5B). Neither mammography, intraoperative palpation, nor histologic analysis revealed a physical connection between these tumors. Five tumors from separate individuals with the same clinical characteristics were also selected as comparisons (Table 1, and Supplementary Table 1).

RNA sample preparation

Total RNA was extracted from dissected tumor containing 90% or greater IDC using TRIzol® (Life Technologies, Grand Island, NY, USA) followed by RNeasy® (Qiagen, Valencia, CA, USA). Total RNA (20 µg) was used to synthesize double-stranded cDNA using Superscript® Choice System (Life Technologies), and the template for an *in vitro* transcription reaction was used to synthesize biotin-labeled antisense cRNA (BioArray™ High Yield RNA Transcript Labeling Kit; Enzo Diagnostics, Farmingdale, NY, USA). Labeled cRNA was fragmented, hybridized, and scanned as described in the Affymetrix GeneChip® protocol (Affymetrix, Inc). Affymetrix U95A arrays contain 12,500 known human genes (see Lipshutz *et al* [3]).

Data analysis

Expression profiles were analyzed using GeneChip® Analysis Suite 3.3 (Affymetrix, Inc). Only relative changes equal to or greater than twofold level of gene expression were considered. The average difference values obtained for each tumor were plotted in log scale on scatter graphs using GeneChip®. The average difference values were exported into JMPIN® 3.2.6 (SAS Institute, Inc). Then, all 12,500 data points were used to generate a pairwise (Pearson's) correlation and, after sorting the average difference values by magnitude, a ranked (Spearman's) correlation coefficient for each set of sample comparisons was calculated [4].

Results

Chip-to-chip variation

In order to measure the variation in gene expression between individual Affymetrix oligonucleotide arrays, the labeled cRNA target prepared from tumors 5A and 5B were each hybridized to two different U95A chips. For each cRNA sample, approximately 50% of the 12,500 transcripts represented on the chip were considered 'present' using Affymetrix algorithms.

The data from the 5A duplicate experiments (5A and 5ADUP) had a pairwise correlation coefficient of 0.995 (Fig. 1). Fifty genes showed a twofold or greater difference in expression between duplicate data sets. Likewise, the 5B duplicate experiments (5B and 5BDUP) had a pairwise correlation coefficient of 0.995, with 36 genes showing a

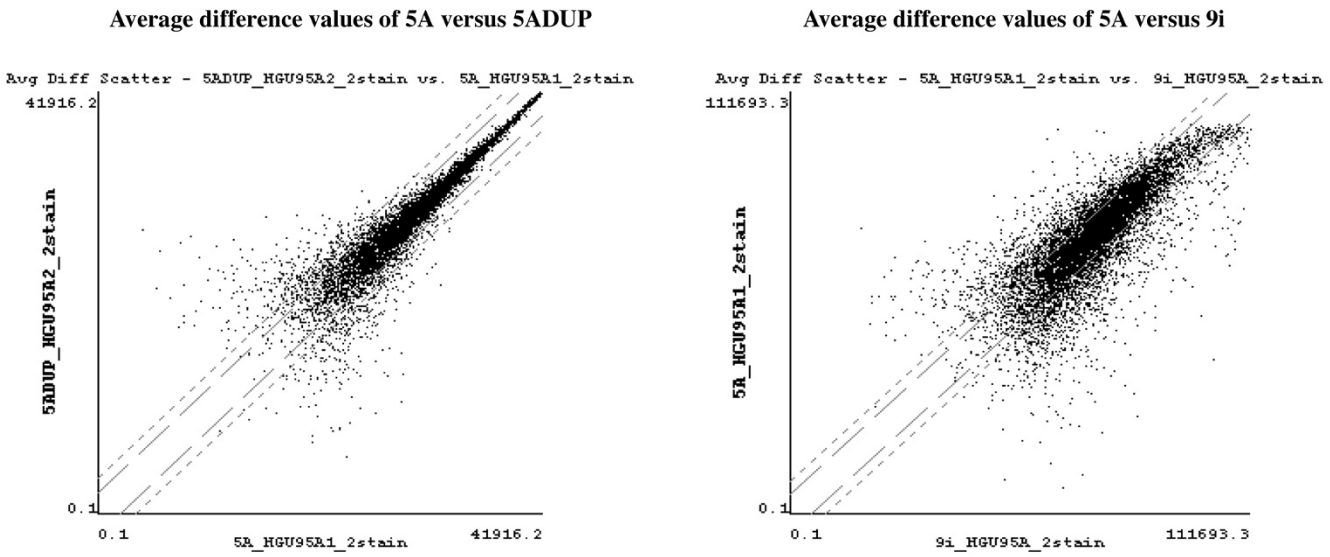
Table 1

Tumor diagnostic characteristics

	Tumor*			
	5A	5B	7	9
Diagnosis	IDC	IDC	IDC	IDC
Size (cm)	3.5	3.0	2.2	1.5
Grade				
Histologic	2	2	2	2
Nuclear	2	2	2	2
ER status (fmols/mg)	178 (positive)	428 (positive)	250 (positive)	475 (positive)
PR status (fmols/mg)	0 (negative)	4 (negative)	150 (positive)	350 (positive)
S phase	3%	2%	Ill-defined G_0/G_1 peak	4%
Ploidy	Diploid	Diploid	Diploid	Diploid
DNA index	1.0	1.0	1.0	1.0
Comments	No adjacent DCIS	No adjacent DCIS	Adjacent DCIS	Adjacent DCIS

*Tumors 5A and 5B are separate adjacent tumors from the right breast of an 87-year-old woman; tumors 7 and 9 are from a 36- and a 76-year-old woman, respectively. DCIS, ductal carcinoma *in situ*.

Figure 1



Representative log scale scatter plots of average difference values.

twofold or greater difference in expression (Table 2). None of the 36 genes that were differentially expressed in the 5B/5BDUP experiment were also differentially expressed in the 5A/5ADUP experiment, suggesting that these changes represent random experimental variation.

Variables in sample preparation

In order to measure the variation in gene expression seen when different cRNA targets are prepared from the same total RNA source and hybridized to separate U95A chips, total RNA from tumor 5A was used to prepare cRNA three weeks after the initial preparation (5AsRNA). Gene expression patterns from 5A and 5AsRNA had a pairwise correlation coefficient of 0.915, with 165 genes having a twofold or greater difference in expression (Table 2). Four genes from the 5A/5ADUP experiment were also differentially expressed at twofold or greater levels in the comparison of 5A with 5AsRNA. Thus, use of the same RNA source resulted in the additional perturbation of approximately 100 genes over the observed chip-to-chip variation, again most likely representing primarily random experimental variation.

Comparison of adjacent tumor gene expression

The gene expression profiles from tumors 5A and 5B had a pairwise correlation of 0.987, with 149 genes having a twofold or greater difference in expression (Table 2). When the expression profiles of cRNA targets from all duplicate experiments were compared (Table 2), the observed pairwise correlation ranged from 0.982 to 0.987, with 148–182 genes expressed with twofold or greater differences. Thus, the number of differentially expressed genes observed when comparing cRNA targets prepared from tumors 5A and 5B (149–182 genes) was not greater than that observed with multiple sample preparations from the same total RNA source (165 genes). In fact, the pairwise correlation coefficient of the adjacent tumors (0.987) was higher than those observed with multiple cRNAs from the same total RNA source (0.915). Therefore, we conclude that the gene expression profiles of tumors 5A and 5B are so similar that they are indistinguishable by the oligonucleotide microarray analysis preformed here. These findings suggest that tumors 5A and 5B represent multifocal components of the same tumor, despite a lack of apparent physical contiguity.

Comparison of tumors among different individuals

Tumors 5A and 7 had a pairwise correlation coefficient of 0.787, with 1937 genes having a twofold or greater difference in expression (Table 2). Likewise, the expression profiles of tumors 5A and 9 generated a pairwise correlation coefficient of 0.745, with twofold or greater differences in the expression of 1764 genes (Table 2). As expected from the pairwise correlation data, the expression patterns based on data from the 5A and 5ADUP experiments have a closer linear relationship on the scatter plot than that observed between tumors 5A and 9. Interestingly, tumors 7 and 9 differed by only 1158 genes with a pairwise correlation coefficient of 0.874, indicating that these tumors may be more similar to each other than they are to tumors 5A and 5B.

In order to confirm our results in a larger sample set, we preformed an analysis on an additional set of invasive breast tumors from different individuals (tumors 3, 20, and 32) and compared the data with those from tumor 5A, and

Table 2

Correlation coefficient calculations and fold change differences between tumors

Comparison	Pairwise correlation (all data)	Ranked correlations (all data)	Twofold change*	Threefold change*	Fourfold change*
5A–5ADUP	0.995	0.962	50	14	3
5B–5BDUP	0.995	0.968	36	10	1
5A–5AsRNA	0.915	0.936	165	34	6
5A–5B	0.987	0.957	149	64	38
5A–5BDUP	0.981	0.951	174	80	38
5ADUP–5B	0.987	0.952	182	88	48
5ADUP–5BDUP	0.982	0.948	182	83	47
5A–7	0.787	0.854	1937	992	598
5A–9	0.745	0.843	1764	887	557
9–7	0.874	0.877	1158	598	339

*Number of genes with twofold, threefold, or fourfold change or greater in expression.

found similar correlation coefficients to those seen with tumors 7 and 9 (Supplementary Table 2).

Ranked correlation values of samples

Analysis of the distribution of the average difference values of each tumor revealed a skewed distribution (data not shown). Therefore, a nonparametric analysis was used to examine the complexity of the data. Ranked correlation coefficients were generated to match each of the pairwise correlation coefficients calculated (Table 2). A small drop in correlation values (by 0.027–0.035) was observed in the ranked correlations compared with the pairwise correlations in the duplicate experiments and those that compared tumors 5A and 5B (Table 2). Conversely, in the pairwise data sets with lower correlation coefficients (5A/5AsRNA, 5A/7i and 5A/9i), the ranked correlation calculations increased the correlation values by 0.021–0.098. These data suggest that a simple pairwise correlation comparison may not fully represent the relationship between gene expression profiles. Further analyses of these data distributions and algorithmic methods for the evaluation of background and systematic error are currently being conducted.

Expression patterns for selected genes of known interest in breast cancer

Differences in gene expression for specific genes that are implicated in breast carcinogenesis were examined. A breast cancer-associated gene list was derived by querying the Affymetrix EASI® Database U95A, with 'breast' as a search term. A total of 39 genes were identified from this search, but only 22 of these genes were judged present in at least one tumor (Table 3). Tumors 5A and 5B only differed in one breast cancer related gene, the human mammaglobin β precursor gene, and showed no differences in expression for the remaining 22 genes examined, including hormone response, tumor suppressor and mammary specific genes. The human mammaglobin protein is overexpressed in a variety of breast carcinomas [5,6]. It does not appear to be consistently found in a specific subclass of tumors, or is its exact function known. Tumor 5A had significantly higher levels of expression of the vascular endothelial growth factor (*VEGF*) gene (2.0-fold), the breast epithelial antigen (*BA46*) gene (4.3-fold) and the human mammaglobin gene (3.6-fold), as compared with tumor 7.

Tumor 5A had a higher level of estrogen receptor (*ER*) expression than did tumor 7 (12.7-fold), *BARD-1* (2.9-fold), and thymidylate synthetase (3.0-fold). Tumor 5A had higher expressions than did tumor 9 of the progesterone receptor (*PR*; 2.7-fold), human mammaglobin (36.9-fold), and the human mammaglobin β precursor (29.1-fold), but had lower expressions of the androgen receptor (2.5-fold), *VEGF* (2.2-fold), *BARD-1* (3.5-fold) and thrombospondin-1 (2.2-fold). These findings suggest that, even though these tumors appear similar in stage, grade and ER status, they in fact exhibit important differences in gene expression that may be associated with different clinical behaviors.

Discussion

Microarray technology has tremendous potential in cancer research and diagnostics. The genetic profiles of tumors can be assessed and compared in a global manner, thus improving diagnosis and potentially allowing individualization of treatment strategies. Several studies [7–9] have characterized the gene expression patterns of breast cancers, but the present study is the first to test the sensitivity and capabilities of oligonucleotide microarray analysis in determining tumor clonality. The expression profiles of adjacent breast tumors from one individual, and of tumors with similar diagnostic characteristics (including

grade, stage, and hormone status) from different individuals were compared. Additional experiments were preformed to measure chip-to-chip variation and variation in the probe preparation process. All data sets were subjected to both pairwise and ranked correlation calculations. These analyses suggest that complex data sets, such as those derived from microarray experiments, may require multiple-stage analysis to enhance the ability to detect changes optimally and to infer biologic significance from expression profiles.

In the present study, a 0.5% difference was observed in the expression profiles of duplicate experiments using the Affymetrix U95A chips. When the total RNA was hybridized to two different U95A chips, the data were found to have a 0.915 pairwise correlation. The lower than expected correlation coefficient, which was related to sample preparation variation, may be due to a small amount of RNA degradation. Future experiments will confirm whether this variation in sample preparation is predictable as a baseline level. The 5AsRNA expression profile differed from that of 5A by only 165 genes, which is within the range observed in tumor comparisons with a pairwise correlation coefficient of 0.987–0.980. Additionally, when the correlation data were further stratified by rank, the correlation coefficients of 5A and 5AsRNA more closely resembled ranked correlations of the 5A and 5B experiments (Table 2). Regardless, a small amount of chip-to-chip variation and sample preparation variation is to be expected when comparing microarray experiments.

The number of genes that were differentially expressed in tumor 5A as compared with tumor 5B (149–182 genes) approximates the number of genes that are differentially expressed as a result of variations in sample preparation (165 genes). None of the genes that were differentially expressed between these adjacent tumors had a known role in breast cancer development. These findings lead us to conclude that tumors 5A and 5B are essentially indistinguishable by microarray analysis, suggesting that they are likely to be derived from the same clone. Interestingly, little heterogeneity was detected between tumors 5A and 5B, despite the fact that both tumors had each grown to a size of at least 3 cm.

In contrast, the expression profiles of tumors 7 and 9 showed marked differences from that of tumor 5A (correlation coefficients of 0.787 for tumors 5A and 7, and 0.745 for tumors 5A and 9), with the expression of 1937 (tumors 5A and 7) and 1764 (tumors 5A and 9) genes differing by twofold or greater between tumors. These findings were confirmed in a larger sample set of three additional invasive tumors, providing additional support to our findings. It is interesting to note that pairwise comparisions of tumor 5A with other invasive tumors had the lowest correlations of all pairwise comparisons, indicating that tumors 5A and 5B may be phenotypically distinct from all the other invasive tumors examined in the present study.

Table 3

Genes of known interest in breast cancer

	Tumor		
Gene	5A/5B	5A/7	5A/9
ER	NC	D(12.7)	NC
PR	NC	NC	I(2.7)
Androgen receptor	NC	NC	D(2.5)
Epidermal growth factor receptor	NC	NC	NC
ERBB2	NC	NC	NC
VEGF	NC	I(2.0)	D(2.2)
TP53	NC	NC	NC
Ataxia telangiectasia (*ATM*)	NC	NC	NC
FHIT	NC	NC	NC
BRCA2	NC	NC	NC
RAD50	NC	NC	NC
BARD-1	NC	D(2.9)	D(3.5)
Retinoblastoma-1 (*RB1*)	NC	NC	NC
Amplified in breast cancer (*AIB1*)	NC	NC	NC
Breast cancer transcription factor (ZaBCl)	NC	NC	NC
Thymidylate synthetase	NC	D(3.0)	NC
Multidrug resistance gene (*MDR-1*)	NC	NC	NC
Thrombospondin-1	NC	NC	D(2.2)
KI-67 antigen	NC	NC	NC
Breast epithelial antigen (*BA46*)	NC	I(4.3)	NC
Human mammaglobin	NC	I(3.6)	I(36.9)
Human mammaglobin β precursor	I(40.1)	NC	I(29.1)

In addition to several known cancer genes, this list was derived by quering the Affymetrix EASI database for genes on the U95A, using 'breast' as the search term. D(*x*), an *x*-fold decrease in gene expression in this tumor compared with tumor 5A; I(*x*), an *x*-fold increase in gene expression in this tumor compared to tumor 5A; NC, no change in gene expression in this tumor compared with tumor 5A.

When the gene expression profiles of tumors 5A, 5B, 7, and 9 were compared, a difference in gene expression for several genes that are known to be involved in breast cancer was observed (Table 3). Most interestingly, the expression profiles of the tumors examined closely reflect the hormone receptor status, as measured clinically. For instance, tumor 7 had a lower level of ER transcript than did tumor 5A, which was paralleled by a decreased ER content. Tumor 9 had the highest PR gene expression, as well as the highest PR content. We were not able to detect a difference in expression of PR between tumors

5A and 7, however, despite the fact that tumor 7 was PR positive (150 fmol/mg) and 5A was PR negative, possibly representing an expression difference below the sensitivity of the technology (see below). Lastly, HER2 was measured as 'undetectable' in tumor 9 at diagnosis, but was not measured in any of the other tumors examined in the present study. Therefore, the lack of difference in expression of ERBB2 detected in each of these tumors predicts that tumors 5A, 5B, and 9 were also Her-2 receptor negative.

Since the initial draft of this report was prepared, it was learned from Affymetrix that the probe set corresponding to PR on the version of the U95A chip used in the present study was not 'adequately representative' of the human mRNA and promoter region DNA of the PR, on the basis of recent changes to the annotations in the GenBank and UniGene databases. These findings may present an explanation as to why the gene expression data for the PR in this study did not correlate with the protein expression measurements.

Finally, transcripts from several genes that are known to play a role in breast cancer were not detected in any of the tumors analyzed, presumably due to low level expression. These genes include: *BRCA1*, *EGFR*, *VEGF*, and the breast cancer suppressor element, *CP1*. Thus, there is an important subset of genes that may not be detected using this technology because of low transcript levels, or that may not be regulated at the transcriptional level.

In conclusion, these experiments demonstrate the potential for oligonucleotide-based microarrays to assess tumor clonality and tumor heterogeneity. Although it is currently not cost-effective to array every tumor sample using an Affymetrix chip, the results of these experiments may lead to identification of a set of genetic markers that can be used to generate a smaller diagnostic tumor array for use in the clinical setting in the near future.

Acknowledgement

The authors thank the patients for their participation and Joseph Bolmarcich for his critical review of the manuscript. This work was funded by the Susan G Komen Dissertation Grant Award number DISS-00284 (to MAU) and the Breast Cancer Research Foundation (to BLW).

References

1. Zhang L, Epstein J, Band P, Berean K, Hay J, Cheng X, Rosin MP: **Local tumor recurrence or emergence of a new primary lesion? A molecular analysis.** *J Oral Pathol Med* 1999, **28**:381–384.
2. Kim N-G, Roh JK, Kim J-H, Chung WY, Park CS, Kim H: **Clonality analysis using methylation-specific polymerase chain reaction: a novel method for investigating tumor clonality.** *Lab Invest* 1999, **79**:1727–1729.
3. Lipshutz R, Fodor SPA, Gingeras TR, Lockhart DJ: **High-density synthetic oligonucleotide arrays.** *Nat Genet* 1999, **21**:20–24.
4. Dowdy S, Wearden S: *Statistics for Research*, 2nd edn. New York: Wiley & Sons; 1991.
5. Watson MA, Fleming TP: **Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer.** *Cancer Res* 1996, **56**:860–865.
6. Watson MA, Darrow C, Zimonjic DB, Popescu NC, Fleming TP: **Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to Chromosome 11q13.** *Oncogene* 1998, **16**:817–824.
7. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi OP, Wilfond B, Borg A, Trent J: **Gene expression profiles in hereditary breast cancer.** *N Engl J Med* 2001, **334**:539–548.
8. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D: **Distinctive gene expression patterns in human mammary epithelial cells and breast cancers.** *Proc Natl Acad Sci USA* 1999, **96**:9212–9217.
9. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747–752.

Supplementary material

Patients and method

Study subjects and selection criteria

An 87-year-old woman underwent an excisional breast biopsy in 1996 to remove two tumor masses at the 6 o'clock and 9 o'clock positions in her right breast (tumors 5A and 5B, respectively). Neither mammography, intraoperative palpation, nor histologic analysis revealed a physical connection between these tumors. A portion of the tumor was used for clinical diagnostic purposes and, with the patient's consent, the remainder of the tumor was used for research purposes.

Two tumors from separate consenting individuals were identified that had the same stage, grade, and ER characteristics as tumors 5A and 5B, based on the pathology reports (Table 1). The tumors originated from a 36-year-old woman who underwent a left breast excisional biopsy in 1997 (tumor 7) and from a 76 year-old woman who underwent a right excisional breast biopsy in 1998 (tumor 9). The three additional invasive tumors added to our data analysis originated from excisional breast biopsies from a 53 year-old woman (tumor 3), a 44 year-old woman (tumor 20) and a 54 year-old woman (tumor 32; Supplementary Table 1). At the time of diagnosis, the Her-2 receptor was measured as 60% positive for tumor 3 and found to be undetectable in tumor 9. The status of Her-2 receptor was not measured for any of the other tumors used in the present study.

Tumor handling

After the gross pathologic diagnosis was completed, the remaining portion of each tumor used in the present study was snap frozen in liquid nitrogen and placed in a freezer at –80°C until use.

Tumor analysis

In order to confirm the clinical diagnosis and to identify regions of the tumor specimen that contained 90% or greater IDC, frozen sections were prepared from each

Supplementary Table 1

Tumor diagnostic characteristics

	Tumor		
	3	20	32
Diagnosis	IDC	IDC	IDC
Size (cm)	2.6	1.9	1.5
Grade			
Histologic	3	3	3
Nuclear	3	3	3
ER	0 (neg)	0 (neg)	0 (neg)
PR	0 (neg)	0 (neg)	0 (neg)
S phase	No data	No data	15.3%
Ploidy	Aneuploid	No data	Aneuploid
DNA Index	2.4	No data	1.54
Comments	Adjacent DCIS	Adjacent DCIS	Adjacent DCIS

Tumor 3 is from a 53 year-old woman, and tumors 20 and 32 are from a 44- and a 54-year-old woman, respectively. DCIS, ductal carcinoma *in situ*.

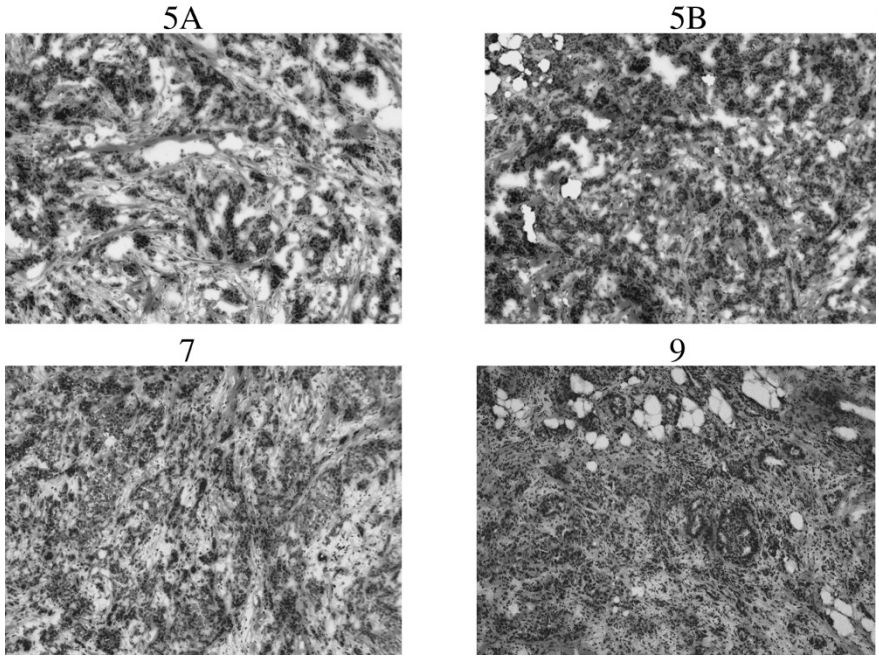
Supplementary Table 2

Gene expression correlations of additional invasive ductal carcinoma tumors

Comparison	Pairwise correlation (all data)	Ranked correlations (all data)
5A–20	0.7880	0.8387
5A–32	0.7104	0.8338
5A–3	0.7228	0.8412
20–32	0.8932	0.8902
20–3	0.8420	0.8490
32–3	0.8594	0.8676
9–3	0.8146	0.8488
9–20	0.8339	0.8266
9–32	0.8786	0.8666
7–3	0.8164	0.8554
7–20	0.8629	0.8362
7–32	0.8516	0.8558

of the tumors. The side of the tumor with the greatest surface area was designated the 'front' of the tumor, and the opposing side was designated the 'back'. A frozen section was made of the front and back of each of the tumor, and the sections were stained with hematoxylin and eosin to visualize the tumor and to verify that the cell population content was consistent throughout the tumor.

Supplementary Figure 1



Hematoxylin and eosin frozen sections taken from tumors. All tumors were diagnosed as IDC, and had similar morphologies and diagnostic characteristics in the clinical pathology report.

Results

Analysis and confirmation of tumor characteristics

Frozen sections stained with hematoxylin and eosin of tumors 5A and 5B contained a 90% or greater IDC population, as determined by a surgical pathologist (MR), and thus did not require further dissection. Tumors 3, 7, 9, 20, and 32 were found to contain small adjacent ductal carcinoma *in situ* components, which were macrodissected from the IDC portion of the tumors. In order to confirm that the ductal carcinoma *in situ* cells were completely removed from the samples, the dissected tumors were resectioned and stained. The frozen sections taken of the 'front' (see Supplementary material, Tumour analysis) of tumors 5A, 5B, 7, and 9 may be seen in Supplementary Fig. 1.