Age-Specific Gene Expression Signatures for Breast Tumors and Cross-Species Conserved Potential Cancer Progression Markers in Young Women

Dilek Colak[1]*, Asmaa Nofal[2], AlBandary AlBakheet[3], Maimoona Nirmal[2], Hatim Jeprel[2], Abdelmoneim Eldali[1], Taher AL-Tweigeri[4], Asma Tulbah[5], Dahish Ajarim[4], Osama Al Malik[6], Mehmet S. Inan[3]†, Namik Kaya[3], Ben H. Park[7], Suad M. Bin Amer[2]*

1 Department of Biostatistics, Epidemiology and Scientific Computing, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 2 Department of Molecular Oncology, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 3 Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 4 Department of Oncology, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 5 Deparment of Pathology and Laboratory Medicine, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 6 Department of Surgery, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia, 7 The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Breast Cancer Research Program, Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America

Abstract

Breast cancer in young women is more aggressive with a poorer prognosis and overall survival compared to older women diagnosed with the disease. Despite recent research, the underlying biology and molecular alterations that drive the aggressive nature of breast tumors associated with breast cancer in young women have yet to be elucidated. In this study, we performed transcriptomic profile and network analyses of breast tumors arising in Middle Eastern women to identify age-specific gene signatures. Moreover, we studied molecular alterations associated with cancer progression in young women using cross-species comparative genomics approach coupled with copy number alterations (CNA) associated with breast cancers from independent studies. We identified 63 genes specific to tumors in young women that showed alterations distinct from two age cohorts of older women. The network analyses revealed potential critical regulatory roles for Myc, PI3K/Akt, NF-κB, and IL-1 in disease characteristics of breast tumors arising in young women. Cross-species comparative genomics analysis of progression from pre-invasive ductal carcinoma *in situ* (DCIS) to invasive ductal carcinoma (IDC) revealed 16 genes with concomitant genomic alterations, *CCNB2, UBE2C, TOP2A, CEP55*, TPX2, *BIRC5, KIAA0101, SHCBP1, UBE2T, PTTG1, NUSAP1, DEPDC1, HELLS, CCNB1*, KIF4A, and *RRM2,* that may be involved in tumorigenesis and in the processes of invasion and progression of disease. Array findings were validated using qRT-PCR, immunohistochemistry, and extensive *in silico* analyses of independently performed microarray datasets. To our knowledge, this study provides the first comprehensive genomic analysis of breast cancer in Middle Eastern women in age-specific cohorts and potential markers for cancer progression in young women. Our data demonstrate that cancer appearing in young women contain distinct biological characteristics and deregulated signaling pathways. Moreover, our integrative genomic and cross-species analysis may provide robust biomarkers for the detection of disease progression in young women, and lead to more effective treatment strategies.

Competing Interests: DC is a PLOS ONE Board Member. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: dcolakkaya@kfshrc.edu.sa (DC); sbinamer2@gmail.com (SMA)

† Deceased.

Introduction

Breast cancer is the most common type of cancer among women worldwide with an estimated 1,300,000 new cases and 465,000 deaths annually [1]. Breast cancer is the major cause of morbidity and mortality among females in Saudi Arabia [2]. Clinical observations indicate that 45% of all female breast cancers in Saudi Arabia developed before the age of 45 years, compared to 9.6% in the United States of America [2,3]. Breast cancer diagnosed in young women is more aggressive in nature with a

poorer prognosis and disease free survival compared to older counterparts [4,5,6,7]. Indeed, it has been shown that survival in younger women is significantly worse for all stages of breast cancer in comparison to older women [8,9]. Although previous studies have described young age is an independent predictor of poor prognosis, the underlying biology driving the aggressive nature of breast cancer arising in young women remains to be elucidated [10,11,12,13].

Typically, the most common histologic type of breast cancer initiates as a premalignant lesion known as atypical ductal

hyperplasia (ADH), then progresses into the preinvasive stage called ductal carcinoma *in situ* (DCIS), and culminates in invasive ductal carcinoma (IDC) [14]. Though it is a multistep process during which genetic alterations accumulate, molecular and pathological evidence suggests that DCIS is a precursor to invasive disease [15,16,17,18]. A genome-wide microarray-based gene expression analysis would be expected to provide an opportunity to discover genes specifically activated or inactivated during the course of breast cancer progression. Despite recent research, the mechanisms underlying tumorigenesis and progression of breast cancer in young women is still not clear [19,20]. In particular, the identification of ''progression markers'' is crucial for determining which lesions are likely to become invasive.

A cross-species comparative genomics approach represents a powerful strategy to identify target genes that may play a role in tumor initiation and progression to malignancy and thus has great therapeutic potential [21,22,23,24,25]. Previous studies have used this approach successfully to understand the molecular pathogenesis of various cancers and disease progression [23,24,26,27]. The rationale is that genomic aberrations and altered pathways involved in oncogenesis are conserved by evolution across different species [24,26,28], and a number of important driver mutations in various cancers have been identified using comparative genomic approaches [24,28,29]. For example, cross-species gene-expression analysis of mouse and human data uncovered gene expression signatures that demonstrate K-Ras oncogene activation in human lung cancers [24]. In another example, Scott Lowe and colleagues identified two oncogenes that are co-amplified and cooperate to promote tumorigenesis by comparing gene amplifications in mouse and human hepatocellular carcinomas [29].

There are areas of genomic instability reported in many cancers, including breast cancer, and some regions commonly exhibit either deletion or increased gene dosage, leading to changes in DNA copy number (CN) [30,31,32,33]. Integrating gene expression with CN data is an effective strategy for interpreting DNA and RNA level anomalies in cancer to identify genes involved with tumor initiation and progression [33,34,35]. Hence, integrating cross-species comparative analysis of human and animal models of breast cancer progression with genomic DNA copy number alterations may lead to robust biomarkers for breast cancer disease progression [32,33,36,37,38].

In this study, we analyzed whole-genome mRNA expression profiling from breast tumors and adjacent normal tissues from Middle Eastern women (n = 113 samples) in age-specific cohorts to characterize the underlying biology of aggressive breast cancers appearing in young women. Moreover, we performed an integrative and cross-species comparative genomics approach to identify evolutionarily conserved marker genes for disease progression in young women and validated its prognostic potential.

Materials and Methods

Patients and Samples

In this study, we focused on breast cancer patients diagnosed with infiltrating ductal carcinoma (IDC) and ductal carcinoma *in situ* (DCIS). Breast cancer samples were collected from primary tumors of 76 patients who sought treatment and underwent surgery (breast conservation surgery or total mastectomy) at the King Faisal Specialist Hospital and Research Center. Signed informed consent was obtained from all patients. On excision of tissues by a surgeon, an anatomic pathologist obtained a sample of the tumor tissue and adjacent normal breast tissue from the same breast having the tumor. 113 samples were collected from patients

and fully consented according to institutional review board approved protocols (KFSHRC IRB Protocol). The study was approved by the research ethics board at our institution (RAC# 2031091). Fresh surgical samples including tumors and adjacent disease free tissues were placed in RNAlater™ (Ambion, Inc) and stored at −20°C after micro dissection had been performed for pathological confirmation. All normal breast tissues were confirmed by the pathologist to have normal morphology before the results were analyzed. Whenever possible depending on the quantity of the surgical samples, a piece of every sample was also snap frozen in liquid nitrogen and then stored at −80°C for subsequent isolation of DNA and proteins. The majority of samples received no prior chemotherapy; only two had chemotherapy and were excluded from further analysis.

Histological assessment of tumors and axillary lymph nodes were done by using formalin-fixed, paraffin-embedded breast cancer samples for HER2, estrogen receptor (ER), and progesterone receptor (PR) status. ER status was determined by immunohistochemistry and measured as a percentage and intensity of positive nuclear staining. The estrogen and progesterone receptors were stained with relevant specific antibodies (Novocastra, Newcastle upon Tyne, UK). For HER2 immunohistochemistry, HercepTest™ (Dako Denmark A/S, Glostrup, Denmark) was used with scores of 0 and 1+ considered negative and 2+ equivocal and 3+ considered positive.

Cancers were categorized as luminal A (ER-positive and/or PR-positive and HER2- and either histologic grade 1 or 2); luminal B (ER-positive and/or PR-positive and HER2+ or ER-positive and/or PR-positive, HER2- and grade 3); HER2 (ER-negative and PR-negative and HER2+); and triple negative (ER-, PR-, and HER2-) as defined previously [39]. Description of the clinicopathological characteristics of patients and breast cancer subtypes for luminal A, luminal B, HER2, and triple negative based on the histological evaluations are shown in Table 1.

Array Hybridization

Total RNA was extracted from tumor and adjacent normal tissue from patients with standard protocols. Sample handling, cDNA synthesis, cRNA labeling and synthesis, hybridization, washing, array (GeneChip® Human Genome U133Plus 2.0 Array, Affymetrix Inc., Santa Clara, CA, USA) scanning, and all related quality controls were performed according to the manufacturer's instructions. The Affymetrix GeneChip/GCOS software (Affymetrix Inc.) was used to calculate the raw expression value of each gene from the scanned image. The total RNA quality was assessed by the values of the $3'-5'$ ratios for actin and glyceraldehyde- 3-phosphate dehydrogenase (GAPDH). DChip [40,41] outlier detection algorithm was used to identify outlier arrays. 104 samples/chips passed the above mentioned quality controls and were used for further analyses. The CEL files were utilized for further analysis using dChip [40,41], MEV [42,43], and PARTEK Genomics Suite (Partek® software, Partek Inc., St. Louis, MO, USA).

Microarray Analysis

Global expression profiling of samples from tumor, IDC (n = 64) and DCIS (n = 7), and adjacent disease free tissues (n = 33) were probed using Affymetrix's GeneChip® Human Genome U133 Plus 2.0 Arrays representing over 47,000 transcripts and variants using more than 54,000 probe sets. The open source R/Bioconductor packages, (Fred Hutchinson Cancer Research Center, Seattle, WA, USA) [44] were employed to normalize the data by the GC Robust Multi-array Average (GC-RMA) algorithm [45,46]. The GC-RMA takes

Table 1. Age-specific patients' characteristics.

Characteristic	All Patients No (%)	Very young (≤35) No (%)	Young (35–45) No (%)	Pre (45–55) No (%)	Old (≥55) No (%)
Type					
IDC	64(90.1)	5(83.3)	24(85.7)	12 (92.3)	23(95.8)
DCIS	7(9.9)	1(16.7)	4(14.3)	1 (7.7)	1(4.2)
Normal	33 (100)	3(9.1)	15(45.5)	7(21.2)	8(24.8)
ER					
positive	49 (69.0)	6(100.0)	15(53.6)	9(69.23)	19(79.2)
Negative	18 (25.4)	0(0)	10(35.7)	4(30.8)	4(16.7)
Missing	4 (5.6)		3(10.7)		1(4.2)
PR					
positive	38(53.5)	5(83.3)	13(46.4)	7(53.9)	13(54.2)
Negative	29(40.9)	1(16.7)	12(42.9)	6(46.1)	10(41.7)
Missing	4(5.6)		3(10.7)		1(4.2)
Grade					
1	3(4.2)	0(0)	0(0)	2(15.4)	1(4.2)
2	38(53.5)	5(83.3)	12(42.9)	7(53.9)	14(58.3)
3	24(33.8)	1(16.7)	12(42.9)	3(23.1)	8(33.3)
missing	6(8.5)	0(0)	4(14.3)	1(7.7)	1(4.2)
HER2					
Positive	33(46.5)	4(66.7)	13(46.4)	6(46.2)	10(41.7)
Negative	34(47.9)	2(33.3)	12(42.9)	7(53.9)	13(54.2)
Missing	4(5.6)		3(10.7)		1(4.2)
Lymph Node					
positive	38(53.5)	2(33.3)	13(46.4)	9(69.2)	14(58.3)
Negative	28(39.4)	3(50.0)	13(46.4)	3(23.1)	9(37.5)
Missing	5(7.0)	1(16.7)	2(7.1)	1(7.7)	1(4.2)
LIVI					
seen	35(49.3)	4(66.7)	10(35.7)	8(61.5)	13(54.2)
absent	31(43.6)	2(33.3)	15(53.6)	4(30.8)	10(41.7)
missing	5(7.0)	0(0)	3(10.7)	1(7.7)	1(4.2)
Subtypes					
Luminal A	24(33.8)	2(33.3)	7(25.0)	6(46.2)	9(37.5)
Luminal B	25(35.2)	4(66.7)	7(25.0)	3(23.1)	11(45.8)
HER2	14(19.7)	0(0)	8(28.6)	4(30.8)	2(8.3)
Triple negative	3(4.2)	0(0)	2(7.1)	0(0)	1(4.2)
Missing	5(7.0)	0(0)	4(14.3)	0(0)	1(4.2)

doi:10.1371/journal.pone.0063204.t001

into account the GC content of the probe sequences when comparing the expression intensities of the different probe sets. To determine significant differences in gene expression levels among different age groups (young women (≤45 years), 45 to 55 years (pre) and ≥55 years (elderly) cohorts), we performed a multi-factor ANOVA including ER, PR, HER2, and grade status as additional factors in a linear additive model, as described previously [47]. We used tumor samples data with complete pathological reports in this model (n = 67). Additionally, we used all tumor and normal samples (n = 104), and performed two-way ANOVA by taking age (young, pre, and elderly), type (tumor or normal), as well as their interaction into the model [47]. In this model, we compared transcriptomes of the tumor tissue and normal tissue for each age group separately. Significantly modulated genes were defined as those with an absolute fold change >2.0 and adjusted p-value <0.05. Multiple hypothesis testing was controlled by applying the Benjamini-Hochberg false discovery rate (FDR) correction. Unsupervised two-dimensional hierarchical clustering using Euclidean distance as well as Pearson's correlation with average linkage clustering was performed. Biological themes associated with the differentially expressed genes was identified by using DAVID Bioinformatics Resources [48], Expression Analysis Systematic Explorer (EASE) [49], and Ingenuity Pathways Analysis (IPA) 6.3 (Ingenuity Systems, Mountain View, CA). Using these bioinformatics tools, we were able to gain greater biological insights into activated or repressed functional processes and altered pathways in the disease pathogenesis compared to the listing of differentially expressed genes. Categorical variables and differences in rates between groups

were analyzed using the χ^2 test. The Fisher exact test was used when expected cell counts were less than 5 using the Monte Carlo method as implemented in SAS. A P-value of <0.05 was considered significant. Statistical analyses were performed by using SAS 9.2 (SAS Institute, Cary, NC), MATLAB (The MathWorks), and PARTEK Genomics Suite softwares. All microarray data reported here are MIAME compliant and have been submitted to the NCBI Gene Expression Omnibus (GEO) database (GSE29044), according to MIAME standards [50].

Independent Datasets

For cross-species analysis, the murine markers of disease progression are taken from Kretschmer et al (Table S4 in [37], GSE21444). Online analysis tools and databases developed by Gyorff et al [51] containing gene expression data and survival information from over 1800 breast cancer patients were obtained and downloaded from Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). These were used to assess the prognostic potential of our gene signature (details of the datasets included in the database are given in the original publication [51]). In addition, The Cancer Genome Atlas (TCGA) data from breast invasive carcinoma (n = 536) and matched normal (n = 63) (https://tcga-data.nci.nih.gov/tcga/), and datasets from GSE7390 [52] and GSE12093 [53] through the canEvolve web portal (www.canevolve.org/) were used for independent validation analyses. The Miller et al. [54] dataset (GSE3494) was also reanalyzed for validation of our gene signature. The GeneSigDB database [55] was used to find the overlap/overrepresentation of our gene signatures with previously published gene signatures for various cancers, including breast cancer. Finally, multiple large genomic data sets with DNA copy number alterations associated with breast cancer were retrieved from the Gene Expression Omnibus database through canEvolve (GSE7545, GSE16619, and GSE9154 data sets) and cBio Cancer Genomics Portals [56] (TCGA, Nature 2012 data [30]) for integrative genomic analysis.

Functional Pathway and Network Analysis

Functional pathway, gene ontology and network analyses were executed using Ingenuity Pathways Analysis (IPA) 6.3 (Ingenuity Systems, Mountain View, CA), a web-delivered application that enables the discovery, visualization, and exploration of molecular interaction networks in gene expression data. The differentially expressed gene lists were mapped to their corresponding gene objects in the Ingenuity pathway knowledge base. These so-called focus genes were then used as a starting point for generating biological networks. A score was assigned to each network in the dataset to estimate the relevance of the network to the uploaded gene list. This score reflects the negative logarithm of the P that indicates the likelihood of the focus genes in a network being found together due to random chance. Using a 99% confidence level, scores of ≥2 were considered significant. A right-tailed Fisher's exact test was used to calculate a p value determining the probability that the biological function (or pathway) assigned to that data set is explained by chance alone.

Realtime RT-PCR Experiments

Confirmatory realtime RT-PCR experiments were performed using the ABI 7500 Sequence Detection System (Applied Biosystems). 50 ng total RNA procured from the same microarray study samples were transcribed into cDNA using a Sensicript Kit (QIAGEN Inc., Valencia, CA, USA) under the following conditions: 25°C for 10 min, 42°C for 2 hrs, and 70°C for 15 min in a total volume of 20 µl. Five differentially expressed genes (*ESR1, IL1RN, SEPP1, TIAM1,* and *SCD*) were selected and primers designed using Primer3 software. After primer optimization, realtime PCR experiments were performed with 6 µl cDNA using Quantitech SyBr Green Kit (QIAGEN), employing GAPDH as the endogenous control gene. All reactions were conducted in triplicates and the data was analyzed using the delta delta C_T method [57,58].

Immunohistochemistry

Validation of protein expression was done using immunohistochemistry. Immunohistochemical staining was performed using standard techniques. Monoclonal anti-TGF-α antibody (Calbiochem, clone 213-4.4, dilution 1:50), monoclonal anti PI3 kinase P85 alpha antibody (Abcam, Cambridge, UK, clone ep380y, dilution 1:20) and polyclonal anti IL1 Receptor I antibody (Abcam, Cambridge, UK, Protein G purified, dilution 1:20) were run manually. Slides were deparaffinized by routine techniques. Antigen retrieval was done in Tris/EDTA buffer, pH 9 heated at 95°C in a microwave for 25 minutes. After blocking endogenous peroxidase activity with a 3% aqueous H_2O_2 solution for 5 minutes, the sections were incubated with primary antibodies overnight at 4°C. Labeling was detected with Envision Plus Detection Kit (Dako, cat. No. K4001). Reaction was detected either by DAB (3, 3-diaminobenzidine, sigma, cat. No. D5905-100TAB) or by AEC (3- amino-9-ethylcarbozale, sigma, cat. No. A-5754). The sections were counterstained with Harris hematoxylin (Acros Organics). Staining was visualized using the DAKO Envision kit according to the instructions of the manufacturer (DAKO, Carpinteria, CA).

Results

Global Expression Profiling in Different Age Cohorts

Genome-wide gene expression profiling provides a comprehensive view of the transcriptional changes that occur during the carcinogenic process and enables the understanding of biology beyond what may be apparent from studies assessing only clinicopathologic features. Here, we first analyzed the whole-genome mRNA expression profile from tumors (n = 71) and adjacent disease free tissues (n = 33) and compared tumor with the normal tissue in each age cohort, young women (≤45 years), 45 to 55 years (pre) and ≥55 years (elderly), separately. We identified 2632, 2029 and 2842 significantly dysregulated genes (up- or down-regulated) present in tumors from young, pre and elderly cohorts (adjusted p value <5% and FC >2), respectively (Figure S1A). To obtain deeper insight into tumor pathogenesis in each age cohort, we performed gene ontology (GO) enrichment and interaction network analyses by using Expression Analysis Systematic Explorer (EASE) [49] and the Ingenuity knowledge base. The network analysis indicated activation of MYC, NF-κB and TGF-β signaling pathways in young, pre and elderly cohorts, respectively (Figure S1B).

Genomic Signature Specific to Tumors Arising in Young Women

We next compared the transcriptomes of tumors across three age cohorts using a multi-factor ANOVA, controlling for ER, PR, HER2, and grade of the tumors (n = 67). The ANOVA identified 567 genes that were significantly modulated among three age groups (unadjusted p<0.01). The unsupervised principal component analysis (PCA) using 567 genes separated samples according to their age group, hence supporting the conclusion that there are distinct gene expression changes associated with tumors that are
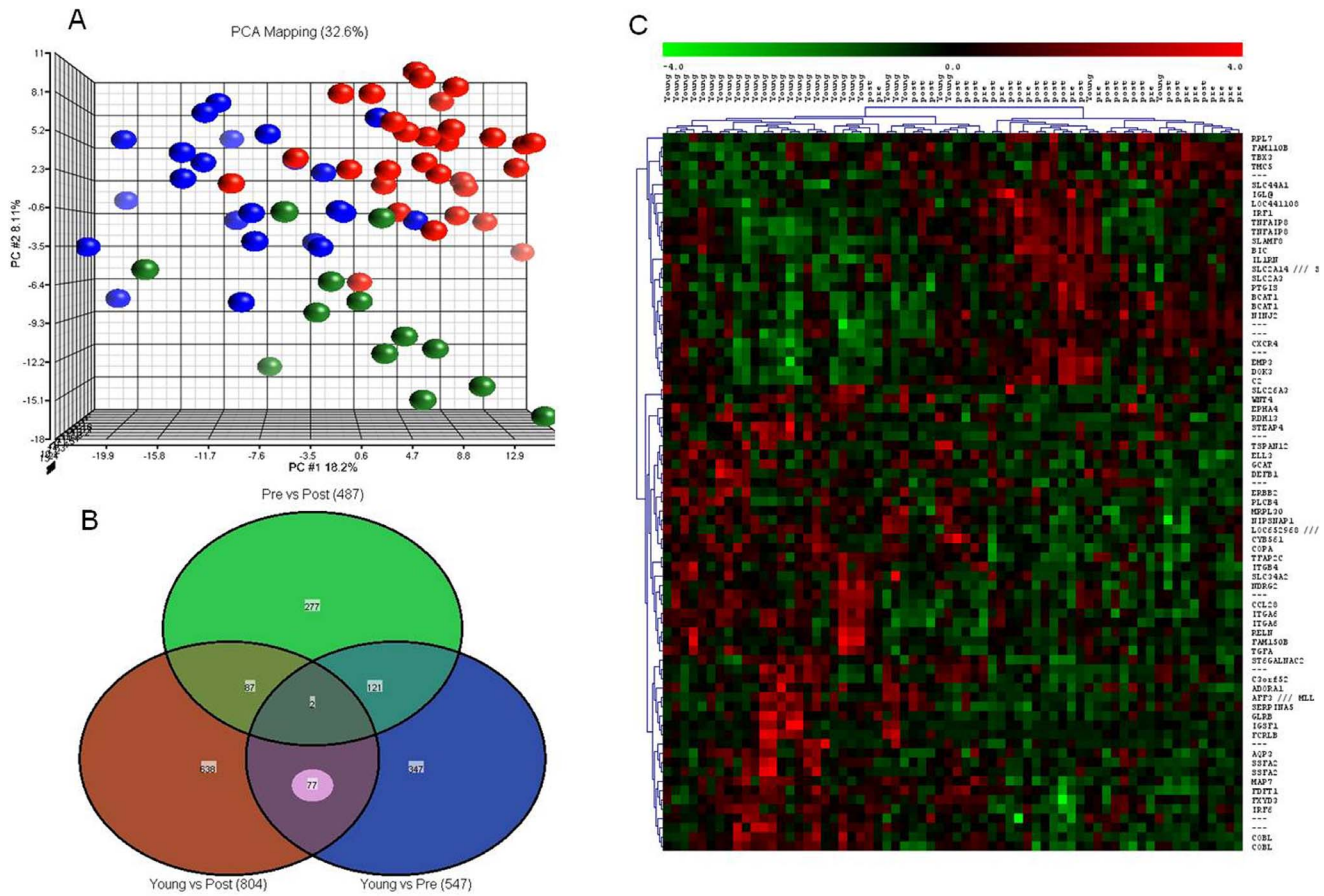
Figure 1. Identification of genes specific to young women with breast cancer. (**A**)The unsupervised principal component analysis (PCA) separated samples according to their age group hence supporting the conclusion that there is a distinct gene expression changes associated with the tumor in different age groups. The red spheres refer to young patients (≤45; Young), green for 45–55 years (Pre), and blue for ≥55 years (Post). (**B**) Venn diagram characterizing differential gene expression between and specific to different age groups. The red circle (left) shows the 804 probes that are differentially expressed between Young and Post; 77 probes (corresponding to 63 genes) were found to be specific to tumor in young women only (circled in *light pink*). (**C**) Unsupervised two-dimensional hierarchical clustering of all tumor samples based on their gene expression similarity using young-age-specific 77 probes was performed using Pearson's correlation with average linkage clustering. The hierarchical clustering revealed clear pattern of genes deregulation defining two main transcriptome clusters, one was mainly composed primarily younger cases, and one was composed of primarily elderly women. Samples are denoted in columns and genes are denoted in rows (gene symbols listed on the right). The expression level of each gene across the samples is scaled to [−4, 4] interval. These mapped expression levels are depicted using a color scale as shown at the bottom of the figure, as such highly expressed genes are indicated in red, intermediate in black, and weakly expressed in green.
doi:10.1371/journal.pone.0063204.g001

dependent on the age of the patient (Figure 1A). We then analyzed overrepresentation of any clinicopathologic or tumor subtype among the age groups, and found no statistically significant associations.

The gene signature specific to tumors in young women (≤45 years) were obtained by overlapping gene lists. When comparing two groups of samples to identify genes differentially expressed in a given group, we used p-value and the fold change (FC) between two groups as the cut-off criteria. As shown in Figure 1B, each circle in the Venn diagram represents the differential expression between two "age groups". This Venn diagram approach revealed that 79 probes were common to both ≤45 vs 45–55 and ≤45 vs >55 comparisons, and 77 probes (corresponding to 63 genes) were specific to tumors in the young group of patients (Y) (shown in pink, in Figure 1B, listed in Table 2) that have significantly higher or lower expression in young women compared to their older counterparts. The unsupervised two-dimensional hierarchical clustering using 63 genes revealed clear patterns of gene deregulation defining two main transcriptome clusters, one was

mainly composed of primarily younger women, and the other one was composed primarily of older patients (Figure 1C). The Microarray Literature-based Annotation (MILANO) database [59] search indicated 98% of those 63 genes had a published association with cancer. Moreover, we tested 63 young age-specific gene signatures against the published gene signatures in Gene-SigDB database [55], and found overrepresentation of our gene set in over 500 gene signatures for various cancers, including breast cancer (adjusted p-value <0.05). The GO and functional analyses revealed significant enrichment of categories, including carcinogenesis, tissue development, cellular development, cellular growth and proliferation, tumor morphology, and cell death (Figure 2A). The network analysis indicated alterations in a number of cancer related pathways, including p38 MAPK, PI3K/AKT, ERK/MAPK and NF-κB signaling pathways, and a potential role of TGFA, ErbB2, and IL-1/IL-1R in young women with breast cancer (Figure 2B).
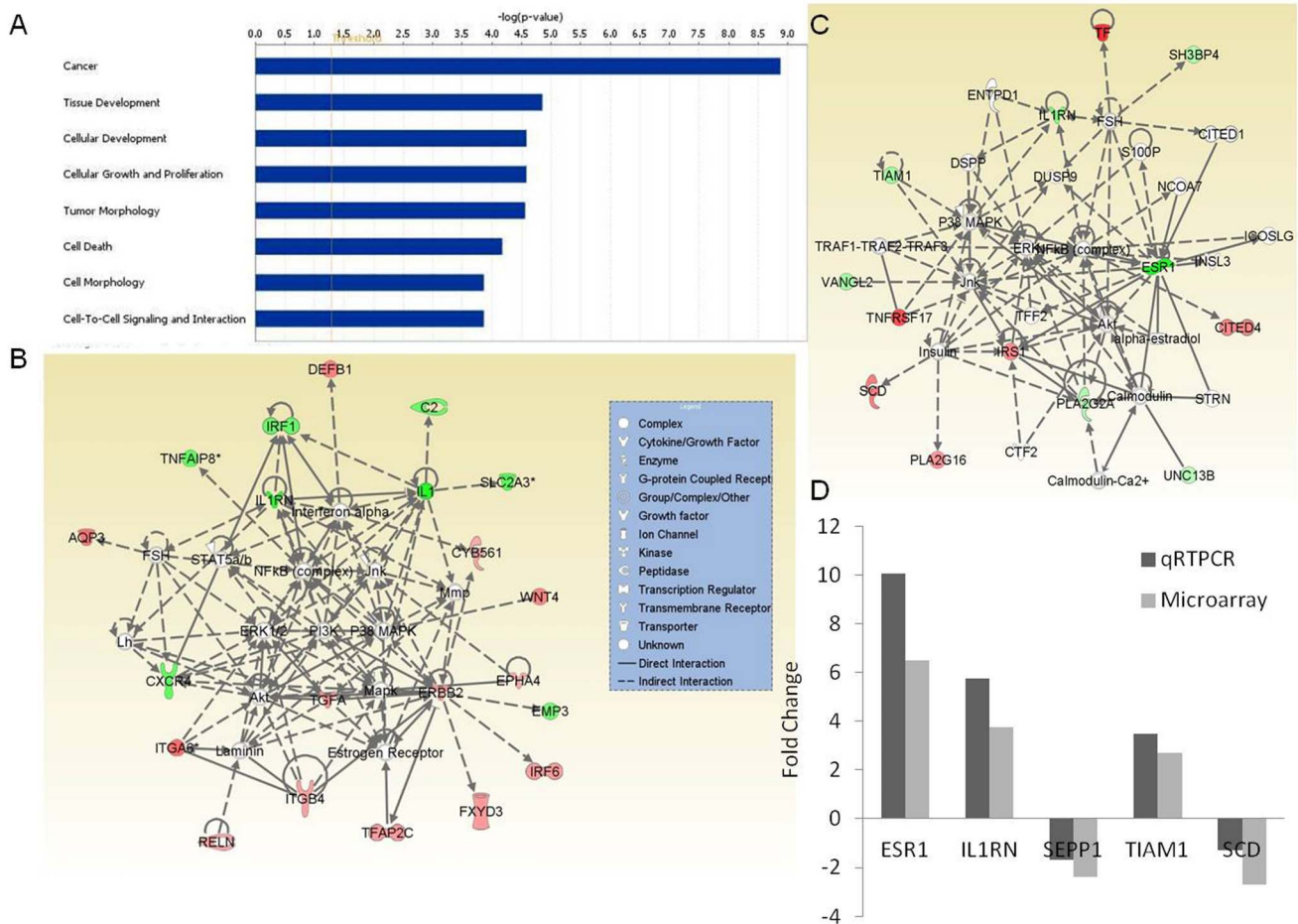
Figure 2. Functional and network analyses of genes specific to young women. (A) The gene ontology and functional analysis of young-age-tumor specific genes (up/down-regulated) were performed using the Ingenuity knowledge base. X-axis indicates the significance (-log P value) of the functional/pathway association that is dependent on the number of genes in a class as well as biologic relevance. The threshold line represents a P value of 0.05. (B–C) Gene interaction network analyses of genes specific to young women and very young women, respectively. Top scoring gene interaction networks with high relevancy scores (with highest relevance score) are shown. Green/red indicates decreased/increased mRNA expression in younger patients compared to older counterparts. The color intensity is correlated with fold change. Straight lines are for direct gene to gene interactions, dashed lines are for indirect ones (D) QRTPCR validation. Grey bars represent microarray hybridizations, and, and dark bars represent values from qRT-PCR. Ratio of expression for each gene in older group (>45) to very young group (≤35) is shown as fold change. A significant correlation existed between the microarray and realtime RT-PCR results.
doi:10.1371/journal.pone.0063204.g002

Genomic Signature Specific to Breast Cancers in Very Young Woman

In Saudi Arabia, almost 50% of all the breast cancer patients were reported to be less than 45 years old. Accordingly, we performed additional analyses within the young women's subset comparing transcriptomes of women younger than 35 years (very young) to two other age cohorts: 35 to 45 years and >45 years. We identified genes that were specific to tumors in very young women using the same methodology that was described previously. The heat map clearly shows significantly higher or lower expression of these genes in very young women compared to the two older age cohorts (Figure S2). The enriched biological processes associated with significantly dysregulated genes that are unique to very young patients include, among others, mitotic cell cycle (p-value = 0.02), morphogenesis (p-value = 0.01), cell proliferation (p-value = 0.03), and death (p-value = 0.049). Similar to young women, network analysis indicated alterations in p38 MAPK, PI3K/AKT and NF-κB signaling pathways, and potentially important roles of IL1RN,

ESR1, and ErbB2 in very young women (Figure 2C and Figure S2).

Cross-Species Comparative Genomics Analysis Coupled with Genomic Alteration Data to Identify Genes that may Play a Role in Cancer Development and Progression in Young Women

Ductal carcinoma in situ (DCIS) is heterogeneous group of pre-invasive tumors which may progress rapidly or slowly to invasive cancer. Therefore, an ability to identify which DCIS lesions are likely to progress to the potentially life threatening stage of invasive ductal carcinoma (IDC) would greatly help in the treatment plan and prognosis of the disease. To identify the putative genes involved in disease progression in young women, we performed genome-wide gene expression profiles characteristic of the sequential disease stages (DCIS and IDC) of breast cancer and compared them to age-matched normal controls in young women (≤45 years). We defined potential progression genes as genes that are significantly altered in both DCIS and IDC as these likely

Table 2. Differentially expressed genes between young women and two older cohorts.

Gene Symbol	Gene Title	FC[a]	FC[b]
Genes with significantly higher expression in the young cohort			
FCRLB	Fc receptor-like B	3.71	3.62
COBL	cordon-bleu homolog (mouse)	3.37	2.91
GLRB	glycine receptor, beta	2.74	2.18
ITGA6	integrin, alpha 6	2.64	2.40
AQP3	aquaporin 3 (Gill blood group)	2.45	3.25
DEFB1	defensin, beta 1	2.33	4.06
SERPINA5	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), membe	2.30	4.24
SLC26A3	solute carrier family 26, member 3	2.27	2.82
WNT4	wingless-type MMTV integration site family, member 4	2.21	2.77
IGSF1	immunoglobulin superfamily, member 1	2.14	2.15
TGFA	transforming growth factor, alpha	2.08	2.75
PLCB4	phospholipase C, beta 4	2.06	2.17
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma de	2.00	2.99
NDRG2	NDRG family member 2	1.96	2.07
MRPL30	mitochondrial ribosomal protein L30	1.93	1.54
AFF3/MLL	AF4/FMR2 family, member 3///myeloid/lymphoid or mixed-lineage leukemia (tritho	1.93	2.19
ST6GALNAC2	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide	1.90	2.32
CCL28	chemokine (C-C motif) ligand 28	1.85	3.23
FAM150B	family with sequence similarity 150, member B	1.84	1.88
FXYD3	FXYD domain containing ion transport regulator 3	1.80	1.99
TFAP2C	transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma)	1.79	1.81
IRF6	interferon regulatory factor 6	1.77	1.88
COPA	coatomer protein complex, subunit alpha	1.77	1.87
ITGB4	integrin, beta 4	1.71	1.83
STEAP4	STEAP family member 4	1.70	1.68
MAP7	microtubule-associated protein 7	1.69	1.53
SLC34A2	solute carrier family 34 (sodium phosphate), member 2	1.68	1.83
RELN	reelin	1.68	1.80
C3orf52	chromosome 3 open reading frame 52	1.68	1.97
RDH13	Retinol dehydrogenase 13 (all-trans/9-cis)	1.67	1.56
ADORA1	adenosine A1 receptor	1.65	2.64
TSPAN12	tetraspanin 12	1.65	2.07
NIPSNAP1	nipsnap homolog 1 (C. elegans)	1.62	1.57
ELL3	elongation factor RNA polymerase II-like 3	1.61	2.12
SSFA2	sperm specific antigen 2	1.60	1.78
EPHA4	EPH receptor A4	1.58	2.22
GCAT	glycine C-acetyltransferase (2-amino-3-ketobutyrate coenzyme A ligase)	1.58	1.95
FDFT1	farnesyl-diphosphate farnesyltransferase 1	1.55	1.62
CYB561	cytochrome b-561	1.55	1.59
Genes with significantly lower expression in the young cohort			
DOK3	docking protein 3	−1.52	−1.54
TBX3	T-box 3	−1.55	−2.21
EMP3	epithelial membrane protein 3	−1.58	−1.64
BIC	BIC transcript	−1.61	−1.75
SLC44A1	solute carrier family 44, member 1	−1.61	−1.72
C2	complement component 2	−1.62	−1.60
LOC441108	hypothetical gene supported by AK128882	−1.63	−1.66
NINJ2	ninjurin 2	−1.64	−2.04

Table 2. Cont.

Gene Symbol	Gene Title	FC[a]	FC[b]
IRF1	interferon regulatory factor 1	−1.64	−1.70
SLAMF8	SLAM family member 8	−1.66	−1.87
RPL7	ribosomal protein L7	−1.68	−1.51
CXCR4	chemokine (C-X-C motif) receptor 4	−1.70	−1.76
PTGIS	prostaglandin I2 (prostacyclin) synthase	−1.72	−2.08
TNFAIP8	tumor necrosis factor, alpha-induced protein 8	−1.73	−2.01
SLC2A14/SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 14///solute	−1.74	−1.70
SLC2A3	solute carrier family 2 (facilitated glucose transporter), member 3	−1.77	−1.73
IGL@	immunoglobulin lambda locus	−1.86	−1.79
BCAT1	branched chain aminotransferase 1, cytosolic	−1.97	−2.30
FAM110B	family with sequence similarity 110, member B	−2.08	−2.82
TMC5	transmembrane channel-like 5	−2.15	−2.70
IL1RN	interleukin 1 receptor antagonist	−2.34	−1.82

[a]FC was calculated between the mean values of expression observed in young women (≤45 years) and ≥55 years.
[b]FC was calculated between the mean values of expression observed in young and 45–55 years.
doi:10.1371/journal.pone.0063204.t002

represent the earliest molecular steps in acquiring the capacity for invasion [15,19,37]. We identified 1015 and 4873 genes differentially expressed (up and down-regulated) in DCIS and IDC compared to normal, respectively, and 697 probes (corresponding to 484 unique genes) that had significantly altered expression in both DCIS and IDC (Figure 3A).

We next performed cross-species comparative genomics analysis to identify potential gene markers for DCIS progression to IDC that are conserved in mouse and human. This approach has been shown to lead to robust markers that may play a role in cancer development and progression. Indeed, driver mutations that are important in cancer have been identified using this strategy [23,24,25,27,29,36]. We used gene expression data from Kretschmer et al. [37] for murine markers of disease progression. The comparison of our progression gene signature with the murine markers (human orthologous) revealed 16 genes that were conserved between mouse and human (p<0.001) (Table 3). GO analyses using both EASE and IPA tools revealed that these genes are mainly involved in biological processes such as cell cycle, mitosis, embryonic development, DNA replication, growth and apoptosis (Figure 3B). The top five significantly altered canonical pathways include Cell cycle: G2/M DNA Checkpoint Regulation (p value = 1.1×10^{-5}), Mitotic Roles of Polo-Like Kinase (p value = 3.3×10^{-5}), ATM Signaling (p value = 1.5×10^{-3}), Cyclins and cell cycle regulation (p value = 2.9×10^{-3}), and Sonic hedgehog signaling (p value = 0.03). The network analysis illustrated activated pathways as well as interactions of genes that may potentially play a role in disease progression (Figure 3C). A literature-based search of 16 genes using the MILANO database [59] demonstrated the association of these genes with cancer progression, tumor development and invasion in various cancers, including breast cancer [22,37,60,61,62,63,64].

The presence of altered DNA CN may contribute to cancer formation and progression and could include transcriptional control mechanisms that locally impact gene expression levels [31,33,65,66]. Integrating the gene expression data with CN alterations may identify novel early breast cancer markers of malignant transformation and progression [33,34,35]. Hence, we integrated our cross-species conserved progression gene signature

with four independent studies of genome copy number alterationsin human breast tumors (as detailed in "Materials and Methods" section) and found that our gene signature has concomitant DNA alterations [30,31,32] (Table 3, Figure 3D).

Comparison of DCIS and IDC Transcriptome in Young Women

Comparison of expression profile characteristics between IDC and DCIS in young women revealed dysregulation of 143 genes, 96% of which had significantly higher expression in DCIS compared to those in IDC (Figure 3A). These genes were enriched within functional categories including immune response, tissue morphology, cellular growth and proliferation, cell death and cellular movement. The network analysis highlighted alterations in PI3K/Akt, NFkB, Jnk, and ERK pathways (Figure S3).

The Venn diagram approach resulted in 27 genes and 94 probes (corresponding to 72 genes) that were unique to IDC and DCIS, respectively (Table S1, Figure S3). Interestingly, 85% of genes specific to IDC were down-regulated compared to normal controls. The IDC gene signature, including *DUSP6, PTGDS, IFNGR1, PIK3R1, FCER1A, P2RY14, PVRL2, SELP*, and TFPI were involved in cell death, immune response, cellular movement and tissue development. The interaction network and pathway analyses revealed alterations in G-Protein Coupled Receptor Signaling, PI3K Signaling, and ERK/MAPK Signaling (Figure S3). In contrast to IDC, 97% of DCIS specific genes were up-regulated in DCIS vs normal, including genes such as *CD22, IGHM, MS4A1, BCR, RBL2*, and *MAP3K5* (Table S1 and Figure S3).

In silico Independent Validations

To validate our results, we used four independently performed microarray datasets as well as data available in the database developed by Gyorffy et al. [51]. The first validation dataset was generated by The Cancer Genome Atlas ((https://tcga-data.nci.nih.gov/tcga/). This dataset is composed of samples from invasive breast carcinoma patients (n = 536) and matched normal controls (n = 63). Our cross-species conserved 16-progression gene signature was significantly up-regulated in patients compared to normal
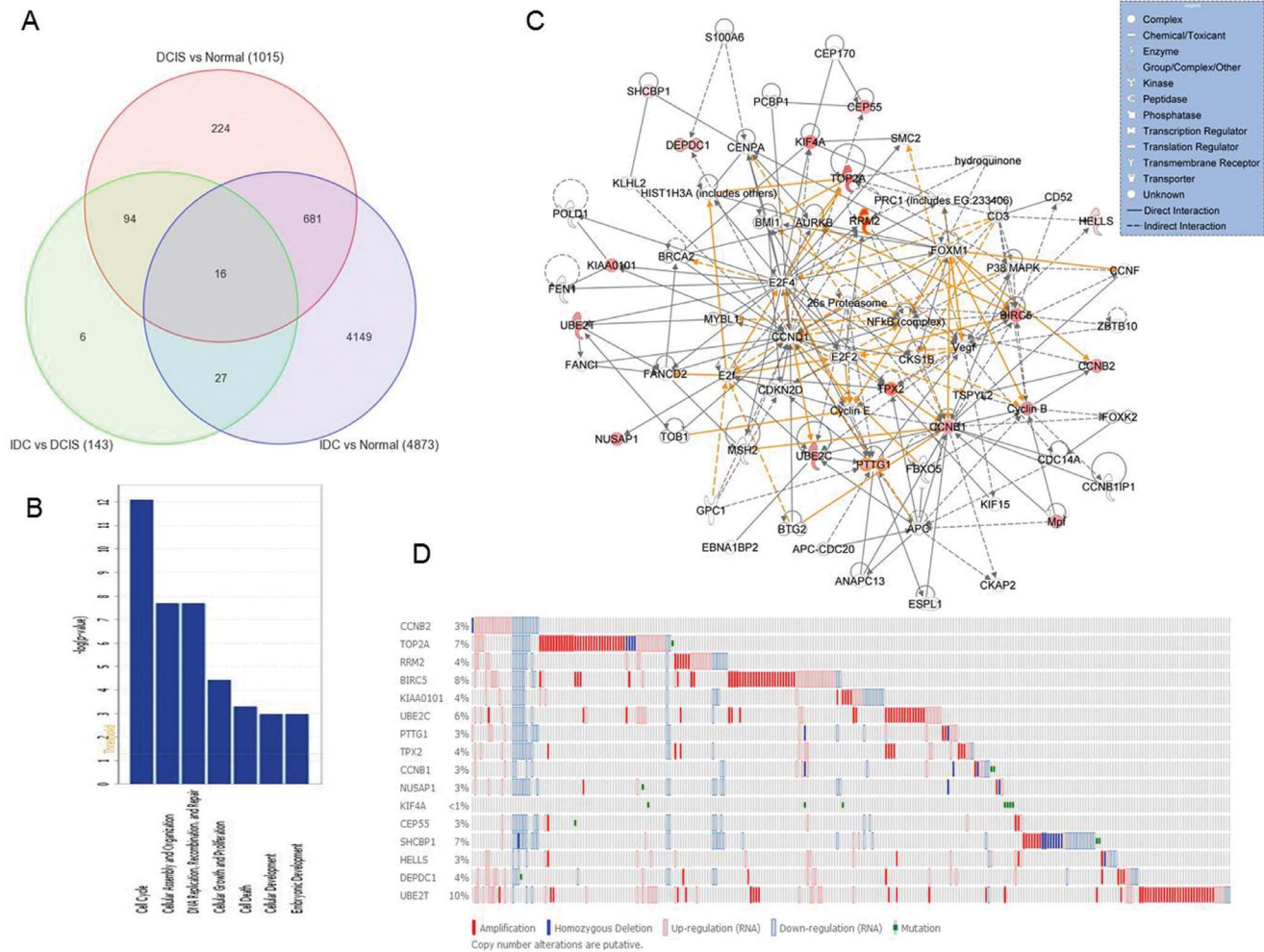
Figure 3. Progression from ductal carcinoma in situ (DCIS) to invasive ductal carcinoma (IDC) in young women. (**A**) The Venn diagram illustrates that there are 1015 genes differentially expressed (up- or down-regulated) in DCIS compared to normal, whereas 4873 genes differentially expressed in IDC compared to normal controls. 143 genes differentially regulated between IDC and DCIS (green circle). (**B**) The functional analysis of 16 potential progression genes identified through cross-species comparative genomics analysis. Y-axis indicates the significance (-log P value) of the functional association that is dependent on the number of genes in a class as well as biologic relevance. The threshold line represents a P value of 0.05. (**C**) Gene interaction networks and pathways analyses of 16-gene progression signature. Green/red indicates decreased/increased mRNA expression in IDC compared to normal controls. The color intensity is correlated with fold change. Straight lines are for direct gene to gene interactions, dashed lines are for indirect ones. (**D**) Invasive breast tumor cases (from TCGA, Nature 2012 [30]) displayed altered amplification/homozygous deletion/up-or down-regulation (RNA) or mutation in our 16-progression gene signature. Cases are denoted in columns, and genes in rows (gene symbols are listed on the left).
doi:10.1371/journal.pone.0063204.g003

controls (adjusted P-value $<1.19\times10^{-32}$) and was sufficient to cluster and differentiate samples as tumor versus normal controls (data not shown).

We then assessed the prognostic capability of our genes on independent microarray datasets involving large numbers of breast cancer patients with survival data. We confirmed the prognostic significance of all of our 16 genes for recurrence free survival (RFS;n = 2324), overall survival (OS;n = 464), and distant metastasis free survival (DMFS; n = 673) in datasets from Gyorffy et al. [51]. The high expressions of these genes were significantly associated with poor disease outcome (Table 3). Moreover, the prognostic significance of 16 genes were tested on additional two datasets of breast cancer patients from GSE7390 [52] and GSE12093 [53]. The GSE7390 dataset consisted of 198 lymph node-negative (N-) patients [52]. The purpose of this analysis was to identify patients at high risk of early distant metastases. The

data from Zhang et al (GSE12093) [53] included 136 breast cancers that were treated with tamoxifen to classify high-risk patients that benefit from adjuvant tamoxifen therapy. We found that thirteen of our genes were significantly associated with a high risk patient group with distant metastases in at least two of the datasets tested (Table 3). Six genes (RRM2, BIRC5, TOP2A, NUSAP1, TPX2, and CCNB2) were of significant clinical relevance in all the datasets tested, especially for identifying a high risk patient group (Table 3, Figure S4).

As a further validation of our results, we re-analyzed an independently performed microarray dataset from Miller et al [54].This dataset was composed of 251 human breast cancer samples, of which 31 were derived from young women, which were used in this re-analysis. We evaluated the performance of the 16-progression gene signature on this dataset. Unsupervised clustering was performed and we found that our gene signature

Table 3. List of 16 cross-species conserved DCIS to IDC potential progression gene signature.

Gene[1]	Gene Name	DCIS[a]	IDC[b]	Biological process term	HR (95%CI)[c]	p-value[d]	GSE7390[†]	GSE12093[‡]
CCNB2*	cyclin B2	4.83	4.76	embryonic development; cell cycle; mitosis; growth	2.23(1.94–2.56)	<1e-16	7.6e-4	0.016
TOP2A*,**	topoisomerase (DNA) II alpha 170kDa	8.38	7.9	DNA metabolic process; DNA replication	2.21(1.92–2.54)	<1e-16	5.8e-3	2.9e-3
RRM2*	ribonucleotide reductase M2 polypeptide	23.96	10.4	DNA replication	2.01(1.75–2.31)	<1e-16	0.01	4.1e-4
BIRC5*	baculoviral IAP repeat-containing 5	7.28	6.10	G2/M transition of mitotic cell cycle; mitosis; cell division	2.07 (1.8–2.38)	<1e-16	0.03	4.7e-3
KIAA0101*	KIAA0101	4.76	5.12	–	1.78 (1.55–2.05)	1.1E-16	NS	7.5e-3
UBE2C*	ubiquitin-conjugating enzyme E2C	4.98	5.83	ubiquitin-dependent protein catabolic process; cell cycle; mitosis; protein ubiquitination	1.81 (1.58–2.08)	<1e-16	0.02	NS
PTTG1*	pituitary tumor-transforming 1	4.23	3.76	DNA metabolic process; DNA repair; cell cycle; mitosis; cell division	1.71 (1.49–1.96)	8.1E-15	NS	2.9e-3
TPX2*	TPX2, microtubule-associated, homolog (Xenopus laevis)	6.95	7.28	apoptosis; cell cycle; mitosis; cell proliferation; cell division	1.83 (1.59–2.1)	<1e-16	0.01	0.014
CCNB1*,**	cyclin B1	5.63	4.85	G2/M transition of mitotic cell cycle; mitotic cell cycle; in utero embryonic development; cell division cycle	2.36 (2.05–2.72)	<1e-16	NS	0.02
NUSAP1*,**	nucleolar and spindle associated protein 1	7.40	4.85	cell cycle; mitosis; cell division	1.82 (1.59–2.09)	<1e-16	0.02	0.04
KIF4A**	kinesin family member 4A	5.41	6.4	organelle organization; microtubule-based movement	1.91 (1.66–2.19)	<1e-16	0.01	NS
CEP55*,**	centrosomal protein 55kDa	7.27	4.7	cell cycle; mitosis; cell division	2.2 (1.92–2.53)	<1e-16	NS	0.04
SHCBP1*,**	SHC SH2-domain binding protein 1	3.72	2.17	–	1.87 (1.63–2.14)	<1e-16	0.048	NS
HELLS*	helicase, lymphoid-specific	4.11	2.05	cell cycle; mitosis; multicellular organismal development; lymphocyte proliferation; cell division	1.74 (1.51–1.99)	1.7e-15	NS	NS
DEPDC1*,**	DEP domain containing 1	5.61	3.50	signal transduction	1.26 (1.1–1.44)	7.5e-4	NS	NS
UBE2T*	ubiquitin-conjugating enzyme E2T (putative)	5.83	5.56	DNA repair, protein monoubiquitination; response to DNA damage stimulus	na	na	na	na

[1]Genes with asterisk are also located in the chromosomal CNA region and ** Mutation found in patients in [30].
[a]DCIS indicates fold change between the mean values of expression observed in DCIS (ductal carcinoma in situ) and age-matched normal controls.
[b]IDC indicates fold change between the mean values of expression observed in IDC (invasive ductal carcinoma) and age-matched normal controls.
[c]Hazard ration (HR) with 95% confidence intervals and [d]logrank P-value for Recurrence free survival (RFS) using data from Gyorffy et al [51].
[†]and [‡] logrank P-value for distant metastasis free survival using data from GSE7390 [52] and GSE12093 [53], respectively. NS. Not significant; na:Not available.
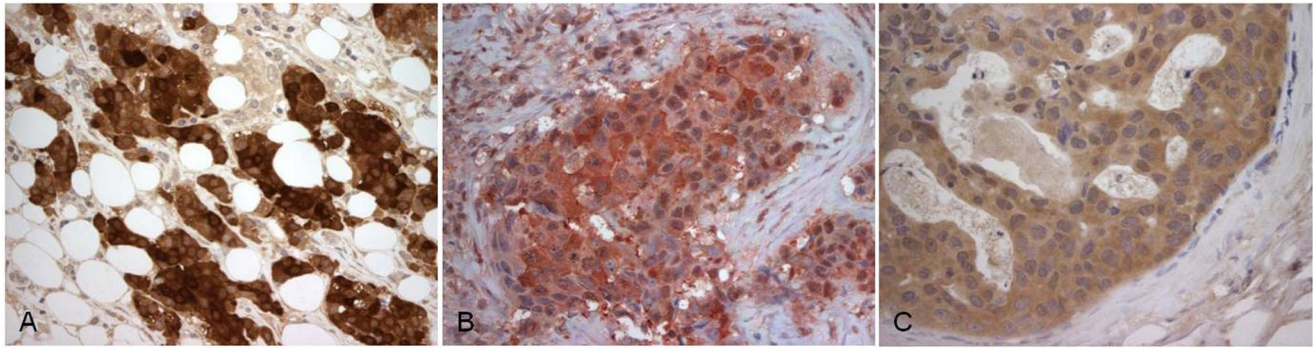doi:10.1371/journal.pone.0063204.t003

Figure 4. Protein expression of selected genes by immunohistochemical staining in breast cancer patients' samples using antibodies directed against (B) TGFA, (C) IL1RN, and (D) PI3K. Representative images of positively stained tumors are shown (magnification, ×200).
doi:10.1371/journal.pone.0063204.g004

was sufficient to separate patients into two clusters which differed significantly by p53 mutation status (Figure S4). The cluster which had high expression of these genes comprised nearly of all the p53 mutant tumors. Intriguingly, TP53 mutations in breast cancer are associated with poor survival independent of other risk factors [67].

The Microarray Literature-based Annotation (MILANO) database [59] search revealed that all of the 16 genes were associated with tumor progression, development, and invasiveness in various cancers, including breast cancer [60,61,62,63]. Moreover, comparing the 16-gene signature with gene signatures available in the GeneSigDB database [55] revealed statistically significant overlap (P-value <0.05, corrected for multiple testing) with over 400 published cancer gene signatures for various cancers, including 161 gene signatures for breast cancer. Furthermore, these genes were also mapped to human genomic CN alterations associated with invasive breast tumors in independent genomic studies, implicating the involvement of these genes in malignant transformation and progression [33,34,35].

Validation of Microarray Data by qRT-PCR and Immunohistochemistry

To confirm the microarray results by an independent method, we selected five significantly dysregulated genes (ESR1, IL1RN, SEPP1, TIAM1, and SCD) in very young (≤35 years) and/or young (≤45 years) women compared to older cohorts and validated the expression levels using qRT-PCR. A significant correlation existed between the microarray and realtime RT-PCR results, (Figure 2D and Figure S2 (Pearson's r >0.76). This correlation was stronger when comparing the older group (>45 years) to the very young women cohort (≤35 years) (r = 0.99; Figure 2D) versus comparing the young woman group (35–45 years) to the very young women cohort (r = 0.77; Figure S2).

Moreover, we performed immunohistochemical staining in breast cancer patient samples using antibodies directed against TGFA, IL1RN and PI3K. The TGFA positivity was significantly associated with young age (Fisher's exact test, p value = 0.02). In fact, 90% of young patients (n = 10) tested positive, which is in concordance with the microarray result. IL1RN was found to have higher expression in older cohorts compared to young patients in our microarray analysis, which was also validated by qRT-PCR (Figure 2D). Indeed, five of the six samples that tested positive by immunohistochemical staining were from older patients. Testing for protein expression of PI3K revealed that it was not expressed in all of the IDC cases (n = 10), but positive for DCIS, which is also

in concordance with the microarray result (Figure S3). Hence, the immunohistochemistry verified the protein expression of the selected candidates. Representative images of positively stained tumors are shown in Figure 4A–C, respectively).

Discussion

Numerous studies have shown that younger women with breast cancer have a poorer prognosis and disease free survival compared to their older counterparts [4,5,6,7,13,68]. Indeed, young age has been shown to be an independent predictor for poor prognosis even after controlling for different histopathological features [13,69]. However, the biology driving this disease process and the molecular pathways that contribute to aggressive tumors in younger women are largely unknown. Clinical observations indicate that 45% of all female breast cancers in Saudi Arabia appear in women younger than 45 years of age [2]. Hence, in this study, we sought to understand the molecular underpinnings of breast cancer in an age-specific manner in order to elucidate genes and pathways giving rise to aggressive tumors in young women using a transcriptomic approach. Furthermore, we explored molecular alterations of breast cancer progression from DCIS to potentially lethal stages of IDC in young women and identified potential progression marker genes using cross-species comparative genomics analysis.

We performed two different approaches to identify gene signatures for different age cohorts of women with breast cancer. In the first approach, we compared whole-genome mRNA expression profile from tumors and disease free normal tissues in three age cohorts of young women (≤45 years), 45 to 55 years (pre) and ≥55 years (elderly). The network analyses of significantly dysregulated genes revealed the activation of MYC [68,70,71], NF-κB [72] and TGF-β signaling [73,74] pathways in young, pre and elderly cohorts, respectively. In the second approach, we compared transcriptomes of tumors arising in young women to those from two older counterparts, and identified 63 genes that had distinct expression patterns in young women. By performing these approaches, we gained important insights into pathways and genes that were specifically altered in young women. The pathway analysis indicated alterations in PI3K/Akt [75,76], MYC [68,70,71] and NF-κB [72] signaling pathways, and potential critical roles for *TGFA* [77,78], *ErbB2* [7,79,80,81], and *IL-1/IL-1R* [82,83,84] which may promote angiogenesis, tumor growth, and metastasis and hence cause the aggressive phenotype observed in young women. Previous reports have shown in experimental models that Interleukin 1 (*IL-1*) promotes angiogenesis, tumor

growth, and metastasis [85], and its presence in some human cancers is associated with aggressive tumor biology [86]. The activation of *IL-1/IL-1R* though autocrine or paracrine mechanisms can lead to a cascade of secondary tumorigenic cytokines, which can subsequently contribute to angiogenesis, tumor-cell proliferation and tumor invasion [82]. For example, these inflammatory cytokines can regulate the proliferation of breast cells through estrogen production by the steroid catalyzing enzymes in breast tissues [87]. Interestingly, mutant alleles of *IL1RN* were associated with shortened disease-free and overall survival among Caucasian women with breast cancer [83]. Similarly, *IL-1* expression has been shown to be an adverse prognostic factor [84,88]. NF-κB signaling has been shown to be activated in various tumors, including human breast cancers. Most recently, it has been shown in mouse models that epithelial NF-κB is an active contributor to tumor progression, inhibition of which could have a significant therapeutic impact even at later stages of mammary tumor progression [89]. Our data also indicated that the levels of expression of *TIAM1* and *VANGL2* in very young women are significantly lower than in their older counterparts. The expression of *TIAM1* has been shown to be associated with increased invasiveness and progression of breast carcinomas [90]. Recently, it has been reported that *VANGL2* promotes migration of cells by a metalloproteinase-dependent invasion of extra cellular matrix and therefore influences invasion and perhaps metastasis [91].

Previous studies have shown that important driver mutations in various cancers can be identified using comparative genomic approaches [21,23,24,28,29]. Such studies suggest that the conserved changes across species may be mechanistically essential for cancer development and progression, and hence they may be critical targets for therapeutic intervention [22,28,92]. Therefore, focusing on differentially expressed genes derived from these comparative approaches along with concomitant altered DNA copy number changes may identify novel early breast cancer markers of malignant transformation and progression [33,34,35]. One of the major contributions of this study is the identification of 16 potential disease progression marker genes, including *CCNB2*, *UBE2C*, *TPX2*, *KIF4A*, *BIRC5*, *NUSAP1*, and *RRM2*, using integrative and cross-species comparative genomics analysis. These genes are related to mitosis, cell cycle, embryonic development, DNA replication, cell division and proliferation. Our findings are consistent with previously performed independent studies of breast cancer progression [15,20,37]. However, the novelty of our results is that genes identified in this study were evolutionarily conserved across species, and along with genomic alterations, and we provide evidence for the potential role of previously reported genes as well as new genes in the progression of young women's breast cancer progression.

Testing our genes on independent microarray datasets using samples from over 3000 breast cancer patients demonstrated that high expression of these genes are significantly associated with poor outcome. Intriguingly, our 16-gene signature separated patients in Miller *et. al.*'s study into two clusters that differed significantly in their TP53 mutation status. The cluster which had high expression of these genes comprised nearly of all the p53 mutant tumors. Previous studies have reported that TP53 mutations in breast cancer are associated with poor survival independent of other risk factors [67] and have a strong association with hormone receptor negative, HER2+ and basal-like subgroups [93,94]. Furthermore, a Microarray Literature-based Annotation database [59] search indicated the involvement of our 16 genes in tumor development, progression, and invasiveness in various cancers, including breast cancer

[22,37,60,61,62,63,64]. Taken together, these observations suggest that the 16-progression-gene signature has the potential to classify tumors which may have invasive capacity and may be crucial for determining which lesions are more likely to become invasive.

Differential expression analysis of DCIS and IDC in young women revealed significant down regulation of PI3K, DUSP6, CD22, RB, BCR, MS4A1 (also known as CD20), and MAP3K5 as well as alterations in PI3K/Akt, NFkB, Jnk, and ERK pathways. The PI3K/Akt pathway is involved in regulation of cell proliferation and implicated in carcinogenesis [95]. The network analysis also indicated a central role of the retinoblastoma tumor suppressor (RB), which may be potentially important in tumor progression. This gene has been found to be functionally inactivated in the majority of human cancers, and aberrant in nearly half of breast cancers [96]. Deficiency in RB function compromises cell cycle checkpoints, and contributes to aggressive tumor proliferation [96]. Comparison of IDC and DCIS transcriptomes resulted in 27 signature genes that are unique to IDC, and differentiated from DCIS in young women. The majority of these genes (85%) were repressed (or down-regulated) compared to normal controls, except for few genes, such as Poliovirus receptor-related 2 (PVRL2, CD112). PVRL2 has been found to have enhanced expression in various tumors, and it has been suggested to have a role in tumor invasion and migration [97,98].

In summary, to our knowledge this study provides the first comprehensive transcriptomic analysis of breast tumors that characterizes the underlying biological mechanisms in an age-specific manner in a cohort of Middle Eastern women, and coupled with an integrative cross-species comparative genomics approach has identified genes that could be potential biomarkers for tumor progression in young women. Our global expression profiling resulted in 63 genes that are specific to young women's breast tumors. The network analyses illustrated the interaction of potential critical genes and the altered pathways associated with breast cancer that specifically appear in young women. The implication from these findings is that these genes may be contributing to the aggressive tumor behavior often present in these patients. Our results confirm previous studies as well as provide additional insights into young age (≤45 years) and very young age (≤35 years) specific oncogenic alterations that may be promoting tumorigenesis. Our cross species data analyses coupled with genomic copy number alterations may provide robust biomarkers for the detection of disease progression in young women and may lead to improved diagnosis and therapeutic options.

Supporting Information

Figure S1 (A) Comparison of each age cohort, young women (≤45 years), 45 to 55 years (pre) and ≥55 years (post), with the age-matched normal controls. We identified 2632, 2029 and 2842 significantly dysregulated genes (up or down) due to tumor in young, pre and old cohorts respectively (adjusted p value <5% and FC >2). (B) Gene interaction networks analysis of differentially expressed genes associated with tumor in each age cohort. Green/red indicates decreased/increased mRNA expression in patients compared to age-matched normal controls. The color intensity is correlated with fold change. Straight lines are for direct gene to gene interactions, dashed lines are for indirect ones (top scoring networks are shown).
(DOCX)

Figure S2 **(A)** Heatmap of very young-specific tumor genes across all tumor samples. Samples are denoted in columns and genes are denoted in rows. The heatmap clearly shows that those set of genes were significantly up- or down-regulated in tumor samples from very young women. The expression level of each gene across the samples is scaled to [-3, 3] interval. These mapped expression levels are depicted using a color scale as shown at the top of the figure, as such highly expressed genes are indicated in red, intermediate in black, and weakly expressed in green. **(B)** Validation of microarray data by realtime RT-PCR. Ratio of expression for each gene in Young (age 35 to 45) to very young (< = 35). Red bars represent microarray hybridizations, and, and blue bars represent values from qRT-PCR. **(C)** Gene interaction networks analysis of genes specific to very young women tumor. Green/red indicates decreased/increased mRNA expression in younger patients compared to older counterparts. The color intensity is correlated with fold change. Straight lines are for direct gene to gene interactions, dashed lines are for indirect ones. (DOCX)

Figure S3 **I.** Comparison of the expression profile characteristics of IDC and DCIS. (a) 143 genes have significantly different levels of expression between DCIS compared to IDC. (b) Functional enrichment analysis of genes whose expression altered between DCIS and IDC. (c-d) The network analysis of 143 genes. Green/red indicates decreased/increased mRNA expression in IDC compared to normal controls. **II.** Network analyses of genes specific to DCIS or IDC in young women (A) Venn diagram illustrating 27 genes and 94 probes (corresponding to 72 genes) that are specific to IDC and DCIS, respectively. (B) Network analyses of genes specific to IDC. Green/red indicates decreased/increased mRNA expression in IDC compared to normal controls. (C) Network analyses of genes specific to DCIS (top two significant networks shown). Green/red indicates decreased/increased mRNA expression in DCIS compared to normal controls. The

color intensity is correlated with fold change. Straight lines are for direct gene to gene interactions, dashed lines are for indirect ones. DCIS: ductal carcinoma in situ; IDC: invasive ductal carcinoma. (PDF)

Figure S4 *In Silico* **Independent Validation Analysis.** (A) Re-analyzed dataset from Miller et al [54] that was composed of 251 human tumor samples, of which 31 were derived from young women, which was used in the re-analysis. Our progression signature gene list was sufficient to separate patients in Miller *et. al.*'s study into two clusters which differed significantly with the p53 mutation status. The cluster which had high expression of these genes comprised nearly of all the p53 mutant tumors. (B) GSE7390 [52] and GSE12093 [53] datasets were used for independent validation analyses. Genes, including RRM2, *BIRC5, TOP2A, NUSAP1,* TPX2, and *CCNB2* were of significant clinical relevance for identifying patients at high risk patients groups (result for *RRM2* has been shown). (DOCX)

Table S1 **Gene signatures specific to malignant stage of invasive ductal carcinoma (IDC) and pre-invasive ductal carcinoma *in situ* (DCIS) in young women.** (DOCX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: SMA DC. Performed the experiments: AN NK AA MN HJ MSI AT. Analyzed the data: DC AE. Contributed reagents/materials/analysis tools: SMA TT DA AE OM NK. Wrote the paper: DC SMA BHP NK.

References

1. Kamangar F, Dores GM, Anderson WF (2006) Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J Clin Oncol 24: 2137–2150.
2. Arabia CRoS (2009) Cancer Incidence Report Saudi Arabia 2005.
3. Society AC (2010) Breast Cancer Facts & Figures 2009–2010. Atlanta: American Cancer Society, Inc.
4. Chung M, Chang HR, Bland KI, Wanebo HJ (1996) Younger women with breast carcinoma have a poorer prognosis than older women. Cancer 77: 97–103.
5. Maggard MA, O'Connell JB, Lane KE, Liu JH, Etzioni DA, et al. (2003) Do young breast cancer patients have worse outcomes? J Surg Res 113: 109–113.
6. Adami HO, Malker B, Holmberg L, Persson I, Stone B (1986) The relation between survival and age at diagnosis in breast cancer. N Engl J Med 315: 559–563.
7. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, et al. (2008) Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol 26: 3324–3330.
8. Bleyer A, Barr R, Hayes-Lattin B, Thomas D, Ellis C, et al. (2008) The distinctive biology of cancer in adolescents and young adults. Nat Rev Cancer 8: 288–298.
9. Nixon AJ, Neuberg D, Hayes DF, Gelman R, Connolly JL, et al. (1994) Relationship of patient age to pathologic features of the tumor and prognosis for patients with stage I or II breast cancer. J Clin Oncol 12: 888–894.
10. El Saghir NS, Seoud M, Khalil MK, Charafeddine M, Salem ZK, et al. (2006) Effects of young age at presentation on survival in breast cancer. BMC Cancer 6: 194.
11. Holli K, Isola J (1997) Effect of age on the survival of breast cancer patients. Eur J Cancer 33: 425–428.
12. Aebi S, Gelber S, Castiglione-Gertsch M, Gelber RD, Collins J, et al. (2000) Is chemotherapy alone adequate for young women with oestrogen-receptor-positive breast cancer? Lancet 355: 1869–1874.
13. Elkum N, Dermime S, Ajarim D, Al-Zahrani A, Alsayed A, et al. (2007) Being 40 or younger is an independent risk factor for relapse in operable breast cancer patients: the Saudi Arabia experience. BMC Cancer 7: 222.
14. Bombonati A, Sgroi DC (2011) The molecular pathology of breast cancer progression. J Pathol 223: 307–317.
15. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, et al. (2003) Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 100: 5974–5979.
16. Chin K, de Solorzano CO, Knowles D, Jones A, Chou W, et al. (2004) In situ analyses of genome instability in breast cancer. Nat Genet 36: 984–988.
17. Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM (2004) Ductal carcinoma in situ of the breast. N Engl J Med 350: 1430–1441.
18. Amari M, Moriya T, Ishida T, Harada Y, Ohnuki K, et al. (2003) Loss of heterozygosity analyses of asynchronous lesions of ductal carcinoma in situ and invasive ductal carcinoma of the human breast. Jpn J Clin Oncol 33: 556–562.
19. Castro NP, Osorio CA, Torres C, Bastos EP, Mourao-Neto M, et al. (2008) Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. Breast Cancer Res 10: R87.
20. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC (2009) Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res 11: R7.
21. Peeper D, Berns A (2006) Cross-species oncogenomics in cancer gene identification. Cell 125: 1230–1233.
22. Gaspar C, Cardoso J, Franken P, Molenaar L, Morreau H, et al. (2008) Cross-species comparison of human and mouse intestinal polyps reveals conserved mechanisms in adenomatous polyposis coli (APC)-driven tumorigenesis. Am J Pathol 172: 1363–1380.
23. Paoloni M, Davis S, Lana S, Withrow S, Sangiorgi L, et al. (2009) Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. BMC Genomics 10: 625.

† Deceased.

24. Sweet-Cordero A, Mukherjee S, Subramanian A, You H, Roix JJ, et al. (2005) An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nat Genet 37: 48–55.

25. Graeber TG, Sawyers CL (2005) Cross-species comparisons of cancer signaling. Nat Genet 37: 7–8.

26. Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, et al. (2003) Myc-driven murine prostate cancer shares molecular features with human prostate tumors. Cancer Cell 4: 223–238.

27. Colak D, Chishti MA, Al-Bakheet AB, Al-Qahtani A, Shoukri MM, et al. (2010) Integrative and comparative genomics analysis of early hepatocellular carcinoma differentiated from liver regeneration in young and old. Mol Cancer 9: 146.

28. Gonzalez-Angulo AM, Hennessy BT, Mills GB (2010) Future of personalized medicine in oncology: a systems biology approach. J Clin Oncol 28: 2777–2783.

29. Zender L, Spector MS, Xue W, Flemming P, Cordon-Cardo C, et al. (2006) Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell 125: 1253–1267.

30. (2012) Comprehensive molecular portraits of human breast tumours. Nature 490: 61–70.

31. Kadota M, Sato M, Duncan B, Ooshima A, Yang HH, et al. (2009) Identification of novel gene amplifications in breast cancer and coexistence of gene amplification with an activating mutation of PIK3CA. Cancer Res 69: 7357–7365.

32. Haverty PM, Fridlyand J, Li L, Getz G, Beroukhim R, et al. (2008) High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer 47: 530–542.

33. Pollack JR, Sorlie T, Perou CM, Rees CA, Jeffrey SS, et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. Proc Natl Acad Sci U S A 99: 12963–12968.

34. Patil MA, Chua MS, Pan KH, Lin R, Lih CJ, et al. (2005) An integrated data analysis approach to characterize genes highly expressed in hepatocellular carcinoma. Oncogene 24: 3737–3747.

35. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, et al. (2005) Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436: 117–122.

36. Ala U, Piro RM, Grassi E, Damasco C, Silengo L, et al. (2008) Prediction of human disease genes by human-mouse conserved coexpression analysis. PLoS Comput Biol 4: e1000043.

37. Kretschmer C, Sterner-Kock A, Siedentopf F, Schoenegg W, Schlag PM, et al. (2011) Identification of early molecular markers for breast cancer. Mol Cancer 10: 15.

38. Bennett CN, Green JE (2008) Unlocking the power of cross-species genomic analyses: identification of evolutionarily conserved breast cancer networks and validation of preclinical models. Breast Cancer Res 10: 213.

39. Collins LC, Marotti JD, Gelber S, Cole K, Ruddy K, et al. (2012) Pathologic features and molecular phenotype by patient age in a large cohort of young women with breast cancer. Breast Cancer Res Treat 131: 1061–1066.

40. Li C, Hung Wong W (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biol 2: RESEARCH0032.

41. Li C, Wong WH (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A 98: 31–36.

42. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, et al. (2006) TM4 microarray software suite. Methods Enzymol 411: 134–193.

43. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374–378.

44. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.

45. Wu Z, Irizarry RA (2004) Preprocessing of oligonucleotide array data. Nat Biotechnol 22: 656–658; author reply 658.

46. Wu Z, Irizarry RA (2005) Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol 12: 882–893.

47. Pavlidis P (2003) Using ANOVA for gene selection from microarray studies of the nervous system. Methods 31: 282–289.

48. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, et al. (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. Nucleic Acids Res 34: W645–650.

49. Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. Genome Biol 4: R70.

50. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 29: 365–371.

51. Gyorffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, et al. (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat 123: 725–731.

52. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, et al. (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13: 3207–3214.

53. Zhang Y, Sieuwerts AM, McGreevy M, Casey G, Cufer T, et al. (2009) The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. Breast Cancer Res Treat 116: 303–309.

54. Miller LD, Smeds J, George J, Vega VB, Vergara L, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 102: 13550–13555.

55. Culhane AC, Schroder MS, Sultana R, Picard SC, Martinelli EN, et al. (2012) GeneSigDB: a manually curated database and resource for analysis of gene expression signatures. Nucleic Acids Res 40: D1060–1066.

56. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, et al. (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2: 401–404.

57. Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19: 368–375.

58. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402–408.

59. Rubinstein R, Simon I (2005) MILANO–custom annotation of microarray results using automatic literature searches. BMC Bioinformatics 6: 12.

60. Psyrri A, Kalogeras KT, Kronenwett R, Wirtz RM, Batistatou A, et al. (2012) Prognostic significance of UBE2C mRNA expression in high-risk early breast cancer. A Hellenic Cooperative Oncology Group (HeCOG) Study. Ann Oncol 23: 1422–1427.

61. Waligorska-Stachura J, Jankowska A, Wasko R, Liebert W, Biczysko M, et al. (2012) Survivin–prognostic tumor biomarker in human neoplasms–review. Ginekol Pol 83: 537–540.

62. Stav D, Bar I, Sandbank J (2007) Usefulness of CDK5RAP3, CCNB2, and RAGE genes for the diagnosis of lung adenocarcinoma. Int J Biol Markers 22: 108–113.

63. Olson JE, Wang X, Goode EL, Pankratz VS, Fredericksen ZS, et al. (2010) Variation in genes required for normal mitosis and risk of breast cancer. Breast Cancer Res Treat 119: 423–430.

64. Chen Z, Zhang C, Wu D, Chen H, Rorick A, et al. (2011) Phospho-MED1-enhanced UBE2C locus looping drives castration-resistant prostate cancer growth. EMBO J 30: 2405–2419.

65. Albertson DG, Collins C, McCormick F, Gray JW (2003) Chromosome aberrations in solid tumors. Nat Genet 34: 369–376.

66. Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, et al. (2005) Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 65: 5561–5570.

67. Pharoah PD, Day NE, Caldas C (1999) Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br J Cancer 80: 1968–1973.

68. Anders CK, Acharya CR, Hsu DS, Broadwater G, Garman K, et al. (2008) Age-specific differences in oncogenic pathway deregulation seen in human breast tumors. PLoS One 3: e1373.

69. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, et al. (2009) Breast cancer in young women: poor survival despite intensive treatment. PLoS One 4: e7695.

70. Xu J, Chen Y, Olopade OI (2010) MYC and Breast Cancer. Genes Cancer 1: 629–640.

71. Corzo C, Corominas JM, Tusquets I, Salido M, Bellet M, et al. (2006) The MYC oncogene in breast cancer progression: from benign epithelium to invasive carcinoma. Cancer Genet Cytogenet 165: 151–156.

72. Park BK, Zhang H, Zeng Q, Dai J, Keller ET, et al. (2007) NF-kappaB in breast cancer cells promotes osteolytic bone metastasis by inducing osteoclastogenesis via GM-CSF. Nat Med 13: 62–69.

73. Buck MB, Knabbe C (2006) TGF-beta signaling in breast cancer. Ann N Y Acad Sci 1089: 119–126.

74. Katsuno Y, Lamouille S, Derynck R (2013) TGF-beta signaling and epithelial-mesenchymal transition in cancer progression. Curr Opin Oncol 25: 76–84.

75. Tokunaga E, Kimura Y, Mashino K, Oki E, Kataoka A, et al. (2006) Activation of PI3K/Akt signaling and hormone resistance in breast cancer. Breast Cancer 13: 137–144.

76. McAuliffe PF, Meric-Bernstam F, Mills GB, Gonzalez-Angulo AM (2010) Deciphering the role of PI3K/Akt/mTOR pathway in breast cancer biology and pathogenesis. Clin Breast Cancer 10 Suppl 3: S59–65.

77. D'Errico A, Barozzi C, Fiorentino M, Carella R, Di Simone M, et al. (2000) Role and new perspectives of transforming growth factor-alpha (TGF-alpha) in adenocarcinoma of the gastro-oesophageal junction. Br J Cancer 82: 865–870.

78. Hantschmann P, Jeschke U, Friese K (2005) TGF-alpha, c-erbB-2 expression and neoangiogenesis in vulvar squamous cell carcinoma. Anticancer Res 25: 1731–1737.

79. Hartley MC, McKinley BP, Rogers EA, Kalbaugh CA, Messich HS, et al. (2006) Differential expression of prognostic factors and effect on survival in young (< or = 40) breast cancer patients: a case-control study. Am Surg 72: 1189–1194; discussion 1194–1185.

80. Agrup M, Stal O, Olsen K, Wingren S (2000) C-erbB-2 overexpression and survival in early onset breast cancer. Breast Cancer Res Treat 63: 23–29.

81. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, et al. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244: 707–712.

82. Pantschenko AG, Pushkar I, Anderson KH, Wang Y, Miller LJ, et al. (2003) The interleukin-1 family of cytokines and receptors in human breast cancer: implications for tumor progression. Int J Oncol 23: 269–284.

83. Grimm C, Kantelhardt E, Heinze G, Polterauer S, Zeillinger R, et al. (2009) The prognostic value of four interleukin-1 gene polymorphisms in Caucasian women with breast cancer: a multicenter study. BMC Cancer 9: 78.

84. Graziano F, Ruzzo A (2005) Role of the interleukin-1 receptor antagonist gene polymorphism (IL-1RN*2) in early gastric cancer. J Clin Oncol 23: 5272; author reply 5272–5273.

85. Saijo Y, Tanaka M, Miki M, Usui K, Suzuki T, et al. (2002) Proinflammatory cytokine IL-1 beta promotes tumor growth of Lewis lung carcinoma by induction of angiogenic factors: in vivo analysis of tumor-stromal interaction. J Immunol 169: 469–475.

86. Elaraj DM, Weinreich DM, Varghese S, Puhlmann M, Hewitt SM, et al. (2006) The role of interleukin 1 in growth and metastasis of human cancer xenografts. Clin Cancer Res 12: 1088–1096.

87. Honma S, Shimodaira K, Shimizu Y, Tsuchiya N, Saito H, et al. (2002) The influence of inflammatory cytokines on estrogen production and cell proliferation in human breast cancer cells. Endocr J 49: 371–377.

88. Jin L, Yuan RQ, Fuchs A, Yao Y, Joseph A, et al. (1997) Expression of interleukin-1beta in human breast carcinoma. Cancer 80: 421–434.

89. Connelly L, Barham W, Onishko HM, Sherrill T, Chodosh LA, et al. (2011) Inhibition of NF-kappa B activity in mammary epithelium increases tumor latency and decreases tumor burden. Oncogene 30: 1402–1412.

90. Stebel A, Brachetti C, Kunkel M, Schmidt M, Fritz G (2009) Progression of breast tumors is accompanied by a decrease in expression of the Rho guanine exchange factor Tiam1. Oncol Rep 21: 217–222.

91. Cantrell VA, Jessen JR (2010) The planar cell polarity protein Van Gogh-Like 2 regulates tumor cell migration and matrix metalloproteinase-dependent invasion. Cancer Lett 287: 54–61.

92. Bennett CN, Green JE (2010) Genomic analyses as a guide to target identification and preclinical testing of mouse models of breast cancer. Toxicol Pathol 38: 88–95.

93. Langerod A, Zhao H, Borgan O, Nesland JM, Bukholm IR, et al. (2007) TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. Breast Cancer Res 9: R30.

94. Rossner P, Jr., Gammon MD, Zhang YJ, Terry MB, Hibshoosh H, et al. (2009) Mutations in p53, p53 protein overexpression and breast cancer survival. J Cell Mol Med 13: 3847–3857.

95. Martelli AM, Cocco L, Capitani S, Miscia S, Papa S, et al. (2007) Nuclear phosphatidylinositol 3,4,5-trisphosphate, phosphatidylinositol 3-kinase, Akt, and PTen: emerging key regulators of anti-apoptotic signaling and carcinogenesis. Eur J Histochem 51 Suppl 1: 125–131.

96. Bosco EE, Knudsen ES (2007) RB in breast cancer: at the crossroads of tumorigenesis and treatment. Cell Cycle 6: 667–671.

97. Sloan KE, Eustace BK, Stewart JK, Zehetmeier C, Torella C, et al. (2004) CD155/PVR plays a key role in cell motility during tumor cell invasion and migration. BMC Cancer 4: 73.

98. Pende D, Spaggiari GM, Marcenaro S, Martini S, Rivera P, et al. (2005) Analysis of the receptor-ligand interactions in the natural killer-mediated lysis of freshly isolated myeloid or lymphoblastic leukemias: evidence for the involvement of the Poliovirus receptor (CD155) and Nectin-2 (CD112). Blood 105: 2066–2073.