

SCIENTIFIC REPORTS



OPEN

Gene expression profiling of calcifications in breast cancer

Sung Ui Shin¹, Jeonghoon Lee², Ju Han Kim², Won Hwa Kim¹, Sung Eun Song¹, Ajung Chu¹, Hoe Suk Kim¹, Wonshik Han³, Han Suk Ryu⁴ & Woo Kyung Moon¹

Received: 14 June 2017

Accepted: 22 August 2017

Published online: 12 September 2017

We investigated the gene expression profiles of calcifications in breast cancer. Gene expression analysis of surgical specimen was performed using Affymetrix GeneChip® Human Gene 2.0 ST arrays in 168 breast cancer patients. The mammographic calcifications were reviewed by three radiologists and classified into three groups according to malignancy probability: breast cancers without suspicious calcifications; breast cancers with low-to-intermediate suspicious calcifications; and breast cancers with highly suspicious calcifications. To identify differentially expressed genes (DEGs) between these three groups, a one-way analysis of variance was performed with post hoc comparisons with Tukey's honest significant difference test. To explore the biological significance of DEGs, we used DAVID for gene ontology analysis and BioLattice for clustering analysis. A total of 2551 genes showed differential expression among the three groups. ERBB2 genes are up-regulated in breast cancers with highly suspicious calcifications (fold change 2.474, $p < 0.001$). Gene ontology analysis revealed that the immune, defense and inflammatory responses were decreased in breast cancers with highly suspicious calcifications compared to breast cancers without suspicious calcifications (p from 10^{-23} to 10^{-8}). The clustering analysis also demonstrated that the immune system is associated with mammographic calcifications ($p < 0.001$). Our study showed calcifications in breast cancers are associated with high levels of mRNA expression of ERBB2 and decreased immune system activity.

Mammography is an established screening tool for breast cancer and calcifications are one of the most important findings for the detection of breast cancer^{1,2}. In addition, the detection and characterization of calcifications are important in preoperative evaluations of lesion extent and surveillance after treatment in breast cancer patients^{3,4}. There are well-established diagnostic criteria based on morphology and distribution of the calcifications at radiologic examination^{5,6}. Evidence suggests that calcifications affect the prognosis of breast cancer, such that breast cancers with calcifications are predicted to be more aggressive and have worse prognosis than those without calcifications⁷⁻¹³. Tabar *et al.* and others have reported that cancers with calcifications of different morphology have different outcomes^{9,14}.

Previous studies have attempted to evaluate and interpret the relationship between mammographic calcifications and tumor histology or the expression of selected biological markers, such as estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2), using immunohistochemistry and have shown that breast cancers with mammographic calcifications are more frequently associated with invasive cancer with extensive ductal carcinoma *in situ* (DCIS) or HER2-positive breast cancer¹⁵⁻²⁰. However, the underlying molecular biology involving calcifications remains poorly understood and needs to be investigated to provide more clues that would help breast specialists make treatment decisions preoperatively⁵.

Linking the relationships between clinical, imaging, pathologic and genomic data is called radiogenomics and there has been growing interest in exploring multiscale relationships in human breast cancers²¹⁻²⁴. In breast cancer, the mainstay of radiogenomic studies often uses magnetic resonance imaging and showed association between several imaging features and genomic data. Identifying the biological background associated with progression and tumor aggressiveness in conjunction with imaging features, such as calcifications, could yield additional data that could assist in pretreatment planning and discussion of prognosis, as well as add to our understanding of tumor biologic characteristics^{18,19}. However, to our knowledge, the relationships between

¹Department of Radiology, Seoul National University Hospital and Seoul National University College of Medicine, Seoul, Korea. ²Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, Korea. ³Department of Surgery, Seoul National University Hospital, Seoul, Korea. ⁴Department of Pathology, Seoul National University Hospital, Seoul, Korea. Correspondence and requests for materials should be addressed to W.K.M. (email: moonwk@snu.ac.kr)

	Breast cancer without suspicious calcifications (n = 99)	Breast cancer with low-to-intermediate suspicious calcifications (n = 37)	Breast cancer with highly suspicious calcifications (n = 32)	p value
Clinicopathologic variable				
Age, years				0.383
≤50	53 (53.5)	16 (43.2)	19 (59.4)	
>50	46 (46.5)	21 (56.8)	13 (40.6)	
Menopausal status				0.191
Premenopausal	51 (51.5)	13 (35.1)	18 (56.3)	
Postmenopausal	48 (48.5)	23 (62.2)	14 (43.8)	
Clinical symptom*				0.151
Yes	91 (91.9)	30 (81.1)	27 (84.4)	
No	8 (8.1)	7 (18.9)	5 (15.6)	
Mean tumor size ± SD (cm) [†]	3.2 ± 1.2	3.1 ± 1.3	2.7 ± 2.1	0.179
Lymph node metastasis				0.255
Yes	43 (43.4)	19 (51.4)	10 (31.3)	
No	56 (56.6)	18 (48.7)	22 (68.8)	
Histologic grade				0.799
I-II	31 (31.3)	10 (27.0)	11 (34.4)	
III	68 (68.7)	27 (73.0)	21 (65.6)	
HR positivity	55 (55.6)	22 (59.5)	15 (46.9)	0.560
HER2 positivity	12 (12.1)	10 (27.0)	16 (50.0)	<0.001
Triple-negative	33 (33.3)	11 (29.8)	5 (15.6)	0.159
DCIS	58 (58.6)	28 (75.7)	31 (96.9)	<0.001
Comedo necrosis	23 (23.2)	21 (56.8)	29 (90.6)	<0.001
Mammographic findings				
Breast composition				0.122
Fatty	34 (34.3)	10 (27.0)	5 (15.6)	
Dense	65 (65.7)	27 (73.0)	27 (84.4)	
Mammographic finding				<0.001
Mass with calcifications	0 (0)	36 (97.3)	28 (87.5)	
Mass without calcifications	88 (88.9)	0 (0)	0 (0)	
Asymmetry or architectural distortion	11 (11.1)	1 (2.7)	4 (12.5)	
Mass shape**				0.003
Oval or round	26 (29.5)	12 (33.3)	0 (0)	
Irregular	62 (70.5)	24 (66.7)	28 (100.0)	
Mass margin**				0.141
Circumscribed	11 (12.5)	4 (11.1)	0 (0)	
Not circumscribed	77 (87.5)	32 (88.9)	28 (100.0)	
Mass density**				0.132
High	77 (87.5)	30 (83.3)	24 (85.7)	
Equal	11 (12.5)	6 (16.7)	4 (14.3)	

Table 1. Patient characteristics according to mammographic calcifications. *HR*, hormone receptor. *HER2*, human epidermal growth factor receptor 2. *DCIS*, ductal carcinoma *in situ*. Unless otherwise indicated, the data are the number of patients, and the data in parentheses are percentages. **p* value was calculated using Fisher's exact. Others were calculated using chi-square test. [†]The numbers represent the mean value ± standard deviation. The *p* value was calculated using ANOVA. **Shape, margin and density of mass were evaluated only in mass cases.

mammographic imaging findings and global transcriptomic profiles are not reported. The purpose of this study was to investigate the gene expression profiles of calcifications in breast cancer.

Results

Patient characteristics. Demographic characteristics and clinicopathologic findings are summarized in Table 1. The median patient age was 50.0 years. The mean tumor size for the study group was 3.1 cm. Patients had invasive ductal carcinomas (89.9%), invasive lobular carcinomas (1.8%) or others with a clinical stage of I (14.3%), II (63.1%), or III (22.6%). Ninety-two patients were hormone receptor (HR) - positive and the other 76 were HR - negative. For HER2 receptor status, 38 patients were positive, and 130 were negative. HER2 positivity, DCIS and comedo necrosis (all *p* < 0.001) were more frequently observed in breast cancers with highly suspicious

calcifications. The other demographic, clinical and pathologic findings were not significantly different between the three groups.

Mammographic features analysis. Mammographic features of the patients are shown in Table 1. The majority of cases presented as mass with or without calcifications ($p < 0.001$). There was no significant difference between the three groups for breast composition or margin and the density of mass. The morphology of calcifications was coarse heterogeneous ($n = 1$), fine pleomorphic ($n = 21$), and fine linear and linear branching ($n = 10$), and the distribution was regional ($n = 1$), grouped ($n = 12$), and segmental ($n = 19$).

Imaging-genomic correlation. When we compared the genomic composition of the three groups, 2551 genes were differentially expressed at the level of $p < 0.05$. Of these, 1838 DEGs (955 up and 883 down) were detected in breast cancers with highly suspicious calcifications compared to those without suspicious calcifications, 484 DEGs (342 up and 142 down) were detected in breast cancers with highly suspicious calcifications compared to those with low-to-intermediate suspicious calcifications, and 457 DEGs (126 up and 331 down) were detected in breast cancers with low-to-intermediate suspicious calcifications compared to those without suspicious calcifications.

The lists of DEGs in all three comparison sets are shown in Supplementary Table 1. With these DEG sets, gene ontology analyses were performed with DAVID. Top 20 genes ordered by p value and fold change in breast cancers with highly suspicious calcifications compared to those without suspicious calcifications are shown in Tables 2 and 3, respectively. There were three genes which are repetitively shown on Tables 2 and 3. ERBB2 gene is up-regulated in breast cancers with highly suspicious calcifications compared to those with low-to-intermediate suspicious calcifications or those without suspicious calcifications. There was no difference in expression levels of ERBB2 between breast cancers with low-to-intermediate suspicious calcifications and those without suspicious calcifications. COL11A1 is down-regulated in breast cancers with highly suspicious calcifications compared to those with low-to-intermediate suspicious calcifications ($p = 0.002$) and those without suspicious calcifications ($p = 1.98E-06$). However, there was no difference in expression levels of COL11A1 between breast cancers with low-to-intermediate suspicious calcifications and those without suspicious calcifications ($p = 0.489$). FNDC1 is up-regulated in breast cancers without suspicious calcifications compared to those with highly suspicious calcifications ($p = 4.67E-06$) and low-to-intermediate suspicious calcifications ($p = 0.014$). However, there was no difference between breast cancers with highly suspicious and those with low-to-intermediate suspicious calcifications ($p = 0.131$).

Focusing on down regulated genes, we found that the 10 top-ranked biological functions (p from 10^{-23} to 10^{-8} , Figure 1) included the immune response, antigen processing and presentation, defense response, the regulation of cytokine production, the positive regulation of immune system and response to wounding. In breast cancers with highly suspicious calcifications compared to those with low-to-intermediate suspicious calcifications, gene ontology biological process terms included two dominant functions; skeletal system development and immune response. Focusing on down regulated genes, we found that top-ranked biological functions ($p < 0.05$) included cartilage development, skeletal system development, limb morphogenesis, osteoblast differentiation, ossifications, immune response, leukocyte mediated immunity, and response to wounding. In breast cancers with low-to-intermediate suspicious calcifications compared to those without suspicious calcifications, focusing on down regulated genes, 10 top-ranked biological functions (p from 10^{-6} to 10^{-2}) included antigen processing and presentation, regulation of apoptosis, cytokine mediated signaling pathway, regulation of programmed cell death, and defense response.

BioLattice analysis identified the lattice of concepts constructed with DEGs between breast cancers with highly suspicious calcifications and those without suspicious calcifications with 60 clusters annotated by gene ontology (GO) terms in the biological process category (Figure 2). Only 24 of 60 clusters demonstrated at least one significant GO term(s) ($p < 0.001$). Overall, the dataset showed 125 significant annotations with 106 unique GO terms. Four core concepts (shown as red color) are associated with the immune system, including defense response, immune response, and inflammatory response. With DEGs between breast cancers with highly suspicious calcifications and those with low-to-intermediate suspicious calcifications, lattice of concept was constructed with 17 clusters annotated by GO terms in the biological process category. Four clusters had significant GO terms ($p < 0.001$). It also contained defense response and immune response. In the comparison of breast cancers with low-to-intermediate suspicious calcifications and those without suspicious calcifications, lattice of concept was made with 16 clusters annotated by GO terms in the biological process category. Only 1 cluster demonstrated significant GO terms ($p < 0.001$). It was composed of immune response, response to pest pathogen or parasite, and response to external biotic stimulus.

Analysis of tumor infiltrating lymphocytes. The mean tumor infiltrating lymphocyte (TIL) score was 37.0 ± 29.5 for breast cancers with highly suspicious calcifications, 34.9 ± 30.4 for those with low-to-intermediate suspicious calcifications, and 42.4 ± 32.8 for those without suspicious calcifications. There was no significant difference between three groups ($p = 0.504$) in total 130 patients. Subgroup analysis according to immunohistochemistry results revealed that there was no significant difference between three groups.

When we combine breast cancers with highly suspicious and low-to-intermediate suspicious calcifications into breast cancers with suspicious calcifications ($n = 52$) and then compare the mean TIL score of it with that of breast cancers without suspicious calcifications ($n = 78$), there was no significant difference between two groups (35.9 vs. 42.4 , $p = 0.251$). However, in triple-negative subtype, the mean TIL score was significantly lower in breast cancers with suspicious calcifications than those without suspicious calcifications (41.5 vs. 62.7 , $p = 0.045$). The other subgroups didn't show the statistical significant differences.

NCBI ID	HUGO	Gene Name	p value	Fold change
94103	ORMDL3	ORMDL sphingolipid biosynthesis regulator 3	2.5E-07	1.582
3038	HAS3	hyaluronan synthase 3	5.8E-07	1.304
338557	FFAR4	free fatty acid receptor 4	7.8E-07	0.758
64175	P3H1	prolyl 3-hydroxylase 1	1.4E-06	0.771
1301	COL11A1	collagen, type XI, alpha 1	2.0E-06	0.400
54894	RNF43	ring finger protein 43	2.8E-06	1.603
84624	FNDC1	fibronectin type III domain containing 1	4.7E-06	0.439
6507	SLC1A3	solute carrier family 1 (glial high affinity glutamate transporter), member 3	5.4E-06	0.677
3983	ABLIM1	actin binding LIM protein 1	7.1E-06	1.461
5996	RGS1	regulator of G-protein signaling 1	7.5E-06	0.531
1501	CTNND2	catenin (cadherin-associated protein), delta 2	8.8E-06	1.530
2064	ERBB2	erb-b2 receptor tyrosine kinase 2	9.0E-06	2.474
80896	NPL	N-acetylneuraminate pyruvate lyase (dihydrodipicolinate synthase)	9.5E-06	0.724
5328	PLAU	plasminogen activator, urokinase	9.7E-06	0.583
713	C1QB	complement component 1, q subcomponent, B chain	1.0E-05	0.681
822	CAPG	capping protein (actin filament), gelsolin-like	1.1E-05	0.833
1573	CYP2J2	cytochrome P450, family 2, subfamily J, polypeptide 2	1.3E-05	1.738
586	BCAT1	branched chain amino-acid transaminase 1, cytosolic	1.3E-05	0.673
2706	GJB2	gap junction protein beta 2	1.5E-05	0.511
10082	GPC6	glypican 6	1.7E-05	0.596

Table 2. Top 20 genes ordered by *p* value in breast cancer with highly suspicious calcifications compared to those without suspicious calcifications.

Discussion

We searched for gene expression profiles of breast cancers with suspicious calcifications and compared those with gene expression profiles of breast cancers without suspicious calcifications. Gene expression patterns were different according to the status of mammographic calcifications in breast cancer. First, breast cancers with highly suspicious calcifications are associated with high levels of mRNA expression of ERBB2 and decreased expression of COL11A1 and FNDC1. Second, GO and clustering analysis using DAVID and BioLattice revealed that breast cancer patients with highly suspicious calcifications on mammography were highly associated with decreased immune system activity.

In our experiments, ERBB2 is repetitively overexpressed shown on the list of top 20 genes ordered by *p* value or fold change and commercially available gene signatures in PAM50, MammaPrint[®] and OncotypeDX[®] (Tables 2–4). To best of our knowledge, it is the first report that insists relationship between mammographic calcifications and mRNA expression of ERBB2. In addition, it supports the findings of other studies and provides bridging evidence that showed there was an association between HER2 overexpression and calcifications in breast cancer patients^{16, 18, 19, 25, 26}. Yepes *et al.*²⁷ reported that a mass with pleomorphic calcifications on mammography may predict an intermediate to high recurrence score in patients with stage I-II ER-positive, HER2-negative, and lymph node negative invasive breast cancer. Similarly, Chae *et al.*²⁸ also reported that the high risk group assessed by 21-gene recurrence score assays was associated with the presence of calcification in the mass and the absence of calcification in the mass is independent predictors associated with low recurrence score. In our study, breast cancers with suspicious calcifications had low expression of COL11A1 and FNDC1. There is little known about FNDC1 gene. COL11A1 is an extracellular matrix molecule which plays an important role in endochondral ossification²⁹. In addition, there are evidences that COL11A1 overexpression is related with up-regulation of TGF- β 1 and a biomarker indicating activated cancer associated fibroblasts in several epithelial cell origin cancers^{30, 31}. Several articles reported that co-cultures of cancer associated fibroblasts with breast cancer cells increased metastatic ability^{32, 33}. It is known to be associated with tumor aggressiveness, tumor progression, infiltration, metastasis and poor survival in several cancers^{30, 34, 35}. However, Fuentes-Martínez *et al.*³⁶ reported that COL11A1 is a stromal marker but does not have prognostic value in breast cancer. To sum up these findings, breast cancer with suspicious calcifications would have another pathway for calcifications formation rather than endochondral ossification and would have little association with stromal remodeling.

To the best of knowledge, it is the first report that breast cancers with mammographic calcifications are associated with decreased immune system activity. Although the direct cellular mechanisms or biologic pathways between calcifications and the immune system have not yet been discovered, we could find one possible explanation. Tse *et al.*³⁷ reported that rapidly proliferating tumor cells that consume the bloody supply result in tumor necrosis and subsequent acidosis in the microenvironment, which finally causes calcium accumulation in the ducts. We can assume that activated immune system may deter the proliferation of tumor cells and necrosis caused by hypoxia^{38, 39}. By contrast, breast cancers with decreased antitumor immune response would have uncontrolled tumor cell proliferation and tumor necrosis, finally causing calcifications in the ducts. Therefore, it is possible that breast cancers with suspicious calcifications are associated with decreased immune system activity. However, there is a conflict within our results showing that breast cancers with suspicious calcifications are

NCBI ID	HUGO	Gene Name	p value	Fold change
118430	MUCL1	mucin-like 1	8.2E-04	12.962
7021	TFAP2B	transcription factor AP-2 beta (activating enhancer binding protein 2 beta)	2.5E-02	5.511
4680	CEACAM6	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen)	3.8E-02	4.977
730	C7	complement component 7	4.8E-04	4.629
219970	GLYATL2	glycine-N-acyltransferase-like 2	6.1E-02	4.263
9635	CLCA2	chloride channel accessory 2	9.4E-05	4.174
339479	BRINP3	bone morphogenetic protein/retinoic acid inducible neural-specific 3	4.0E-04	4.088
79983	POF1B	premature ovarian failure, 1B	8.8E-05	3.583
4322	MMP13	matrix metalloproteinase 13	3.1E-04	0.320
85320	ABCC11	ATP-binding cassette, sub-family C (CFTR/MRP), member 11	2.5E-02	2.776
26154	ABCA12	ATP-binding cassette, sub-family A (ABC1), member 12	8.8E-03	2.666
1300	COL10A1	collagen, type X, alpha 1	2.2E-03	0.381
3081	HGD	homogentisate 1,2-dioxygenase	3.6E-04	2.585
646424	SPINK8	serine peptidase inhibitor, Kazal type 8 (putative)	4.1E-02	2.513
1301	COL11A1	collagen, type XI, alpha 1	2.0E-06	0.400
2064	ERBB2	erb-b2 receptor tyrosine kinase 2	9.0E-06	2.474
10351	ABCA8	ATP-binding cassette, sub-family A (ABC1), member 8	1.5E-03	2.436
90865	IL33	interleukin 33	4.6E-05	2.377
84624	FNDC1	fibronectin type III domain containing 1	4.7E-06	0.439
131578	LRRC15	leucine rich repeat containing 15	1.2E-04	0.443

Table 3. Top 20 genes ordered by fold change in breast cancer with highly suspicious calcifications compared to those without suspicious calcifications.

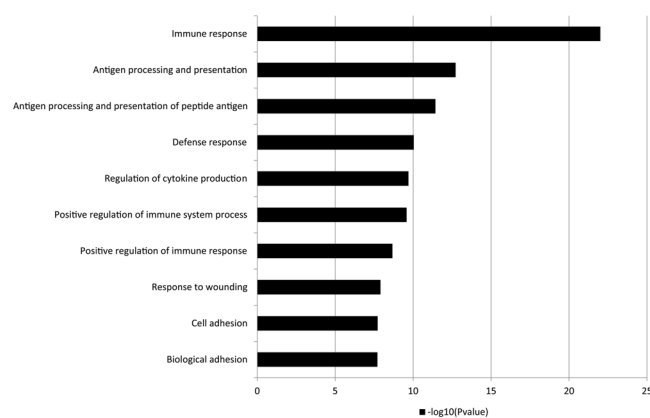


Figure 1. Enriched gene ontology biological process terms using down-regulated DEGs ($n = 882$) from the comparison between breast cancers with highly suspicious calcifications and those without suspicious calcifications. Immune system-related GO terms in the biological process category are enriched in breast cancers without suspicious calcifications.

associated with rapidly proliferating tumor cells following decreased immune system activity. Because DCIS is an indolent non-invasive tumor but it is frequently associated with mammographic calcifications^{40,41}. The mechanisms by which suspicious calcifications are produced may vary between invasive cancer and DCIS and it must be evaluated in future studies. The mean TIL score was significantly lower in breast cancers with suspicious calcifications than breast cancers without suspicious calcifications only in triple-negative subtype. The results of TIL scores partly supports our gene expression analysis that breast cancers with suspicious calcifications are associated with decreased immune system activity. It might suggest that the associations between calcifications and immune system are strong and apparent in triple-negative subtype than others.

Our study has several limitations. First, this study has a retrospective design, and there may be selection bias in our database. Second, we do not have independent validation set to support our results. At the time of our study, the Cancer Imaging Archive of breast TCGA data had only 4 patients with a preoperative mammography. In addition, we couldn't perform analyses regarding survival or recurrence due to lack of long term follow-up data. Thus, we tried to use commercially available gene signatures, however, not all of genes were available on the Affymetrix GeneChip[®] Human Gene 2.0 ST arrays. To overcome the weakness, we analyzed pathologic TIL scores of the same population. However, there remains limitation that we simply hypothesized that TIL might be an indicator of immune system activity, even though immune system is very complex and interactive system.

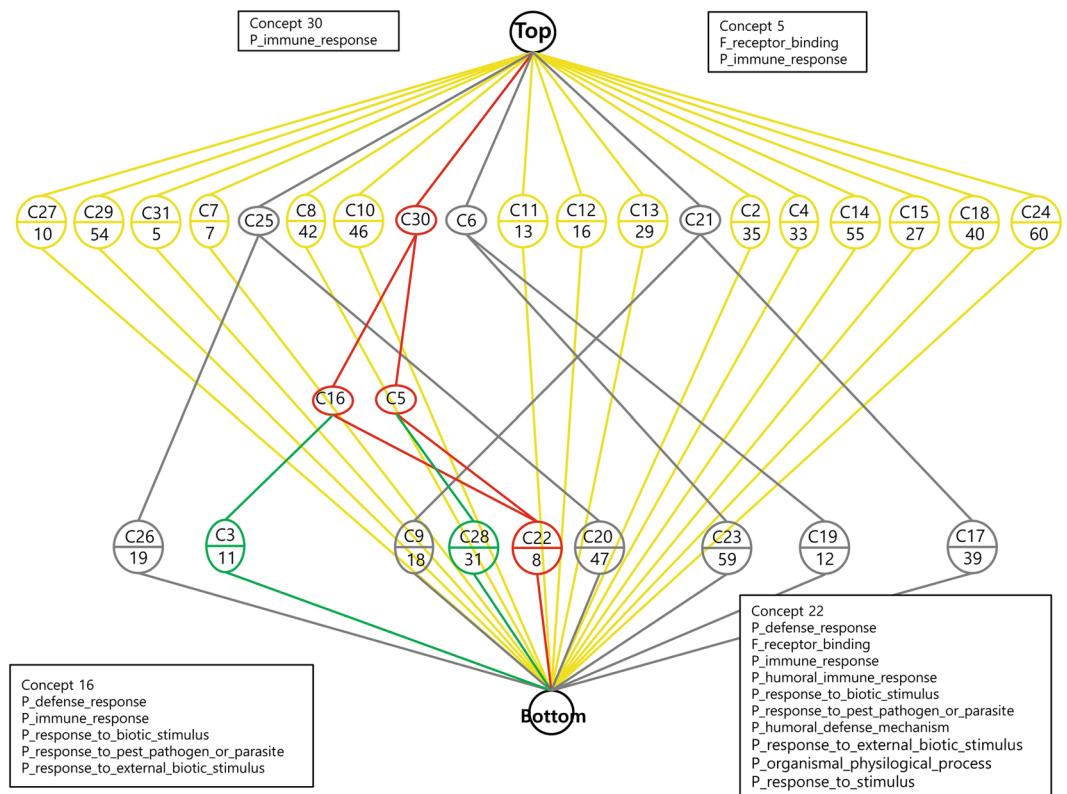


Figure 2. Concept lattice constructed from the comparison between breast cancers with highly suspicious calcifications and those without suspicious calcifications (n = 1838) in a total population with 60 clusters annotated by GO terms in the biological process category. Only 24 of 60 clusters demonstrate at least one significant GO term(s) ($p < 0.001$). Overall, the dataset shows 125 significant annotations with 106 unique GO terms. The core-periphery substructures are marked with colors (i.e., core in red, communicating in green, independent in yellow and peripheral in gray).

NCBI ID	Gene symbol	p value	Fold change		
			Breast cancer with highly suspicious calcifications compared to those without suspicious calcifications	Breast cancer with highly suspicious calcifications compared to those with low-to-intermediate suspicious calcifications	Breast cancer with low-to-intermediate suspicious calcifications compared to those without suspicious calcifications
2064	ERBB2	1.6E-05	2.474	2.254	1.098
2886	GRB7	6.3E-05	2.122	1.875	1.132
968	CD68	1.5E-04	0.727	0.841	0.864
8030	CCDC6	9.2E-04	1.131	1.055	1.072
120224	TMEM45B	9.5E-04	1.859	1.589	1.170
8840	WISP1	2.3E-03	0.798	1.011	0.789
4318	MMP9	3.5E-03	0.728	1.092	0.666
4320	MMP11	5.0E-03	0.682	0.771	0.885
2321	FLT1	7.2E-03	1.183	1.179	1.003
58475	MS4A7	1.0E-02	0.711	0.680	1.045
26996	GPR160	1.0E-02	1.197	1.106	1.082
643008	SMIM5	1.9E-02	1.079	1.057	1.020
3861	KRT14	4.0E-02	1.898	1.623	1.169
2264	FGFR4	4.1E-02	1.298	1.194	1.087
3169	FOXA1	4.8E-02	1.008	0.945	1.067

Table 4. Comparison of commercially available gene signatures in PAM50, MammaPrint® and OncotypeDX®.

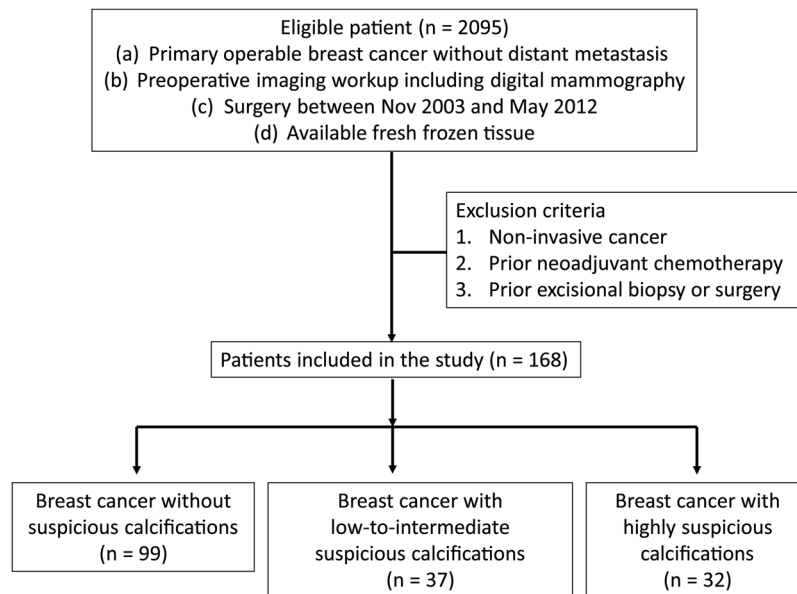


Figure 3. Schematic of the inclusion and exclusion criteria and the selection process of this study cohort.

Third, the interpretation of mammographic calcifications is subjective, and other imaging findings were not considered in this study. Fourth, breast cancer is a heterogeneous disease regarding its gene expression profiles. It is possible that the differences between the groups are not only due to the calcification status but also to the intrinsic subtype of breast cancer. Finally, we did not perform *in vitro* experiments regarding the cellular mechanism of calcium deposit or *ex vivo* experiments to determine whether immune cells differentially exist or whether immune cell markers are differentially presented according to mammographic calcifications. Similarly, we did not perform analysis of specific mineral species in the specimen. However, our study has several distinct strengths. Having both fresh frozen tissues obtained from surgical specimen and initial digital mammography is a rare and valuable resource. Additionally, to the best of our knowledge, this is the first study examining the global gene expression profiles of calcifications in breast cancer, and it is the largest study in which over one hundred patients with microarray data were enrolled and that correlates microarray data and imaging features of breast cancer. In addition, our results may guide further studies in the study of biological process or cell signaling pathways of calcification formation.

In conclusion, gene expression patterns in breast cancer are different according to mammographic calcifications. Breast cancers with highly suspicious calcifications are associated with high levels of mRNA expression of ERBB2 and decreased immune system activity. These results, if validated, could be used as the basis for future hypothesis-based studies.

Methods

Study population. The institutional review board of Seoul National University Hospital (IRB No. 1409-128-612) approved this retrospective study, and all patients provided written informed consent for their breast cancer tissue to be used for genome sequencing (IRB No. 1405-088-580) before operation. All experiments were performed in accordance with relevant guidelines and regulations. We retrospectively identified 2095 consecutive patients with primary operable breast cancer who performed preoperative imaging workup, and underwent surgery between 2003 and 2012 from the Breast Imaging Center database of Seoul National University Hospital. We excluded patients who had (a) non-invasive cancer, (b) prior neoadjuvant chemotherapy; or (c) prior excisional biopsy or breast surgery. As a result, a total of 168 women (mean, 50.7 yrs; age range, 21–79 yrs) comprised our study group (Figure 3).

Mammography acquisition and analysis. Mammography was performed using a Senograph 2000D or Senograph DS (GE Healthcare, Milwaukee, WI, USA) or a LORAD Selenia (Hologic, Boston, MA, USA) digital mammography unit. Standard two-view mammography was performed with additional views as necessary. A Senograph system was used on 104 (61.9%) women, and a Selenia system was used on 64 (38.1%) women.

Mammographic features of the patients were assessed according the Breast Imaging-Reporting and Data System (BI-RADS)⁶. Using calcifications at mammography as a criterion for grouping the patients, three radiologists (S.E.S., A.C., and W.K.M.) with different degrees of experience in interpreting mammography independently analyzed the calcifications without access to genomic data. The radiologists had to fill out a sheet for each case giving their BI-RADS category: 1, normal; 2, benign; 3, probably benign; 4A, low suspicious; 4B, intermediate suspicious; 4C, highly suspicious; and 5, highly suggestive of cancer. After each radiologist finished the analysis, final consensus was established for each case. Cases with BI-RADS category of 4C and 5 were classified as breast cancer with highly suspicious calcifications (n = 32), cases with BI-RADS category of 3, 4A, and 4B as breast

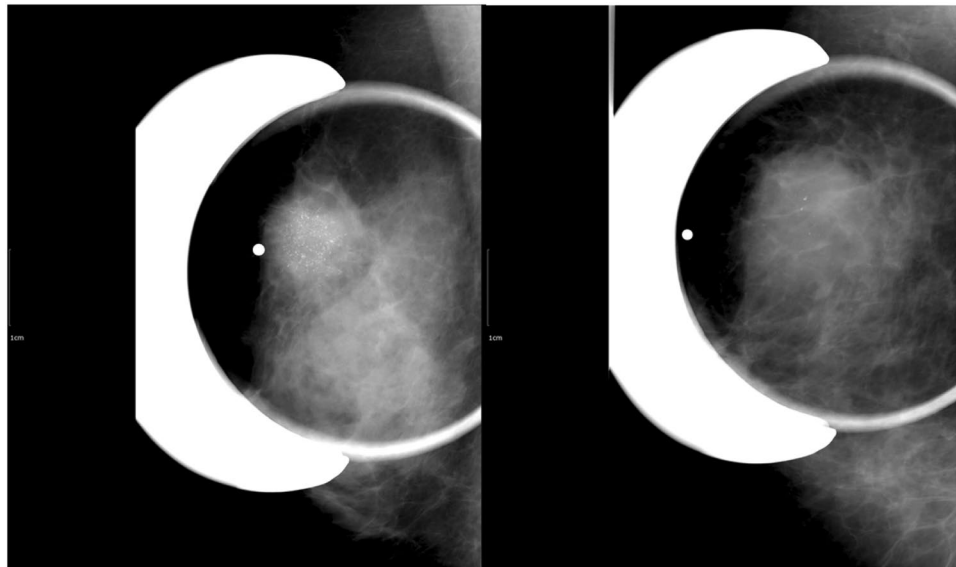


Figure 4. Examples of calcifications. (A) A 55-year-old woman presented with palpable mass in the right breast and magnification mammogram revealed not circumscribed hyperdense mass with grouped fine pleomorphic and fine linear calcifications in the right upper breast. She was categorized as highly suspicious calcifications group and diagnosed as 2.8 cm invasive ductal carcinoma with comedo necrosis. Immunohistochemistry analysis revealed that the tumor was ER-negative, PR-negative, and HER2-positive cancer. (B) A 50-year-old woman presented with palpable mass in the right breast and magnification mammogram revealed not circumscribed hyperdense mass with grouped coarse and amorphous calcifications in the right upper breast. She was categorized as low-to-intermediate calcifications group and diagnosed as 4.5 cm invasive ductal carcinoma. Immunohistochemistry analysis revealed that the tumor was ER-negative, PR-negative, and HER2-negative cancer.

cancer with low-to-intermediate suspicious calcifications ($n = 37$), and the other cases with BI-RADS category 1 or 2 as breast cancer without suspicious calcifications ($n = 99$). For example, fine pleomorphic, fine linear or fine linear branching calcifications were classified as highly suspicious calcifications whereas amorphous or coarse heterogeneous calcifications were classified as low-to-intermediate suspicious calcifications^{6,7} (Figure 4).

Tissue samples and Microarray analysis: RNA Isolation, Preparation, Hybridization, and Data Acquisition. Tissue samples were dissected through the centers of the carcinomatous region during surgery at our hospital between 2003 and 2012. These samples were frozen in liquid nitrogen within 20 min following surgical devascularization and stored at -80°C . Total RNA from each sample was extracted using TRIzol[®] reagent (Invitrogen, Carlsbad, CA, USA). RNA quality was assessed with an Agilent 2100 Bioanalyzer using the RNA 6000 Nano Chip (Agilent Technologies, Amstelveen, The Netherlands), and the quantity was determined with an ND-2000 spectrophotometer (Thermo Inc., DE, USA). The median RNA extracted was 1.202 g/L (range, 0.143–2.986 g/L). Total RNA was measured as the UV absorbance at 260 nm. Sample purity was assessed by measuring the OD 260:280 nm and OD 260:230 nm. The integrity of RNA samples was confirmed by the appearance of distinct 28S and 18S bands of ribosomal RNA. The RNA integrity number (RIN) was determined using the RIN algorithm of the Agilent 2100 Expert Software⁴². The quality of the RNA was good with a standard 260/280 ratio and 260/230 ratio of absorbance greater than 1.7 and 1.3 per sample, respectively. The mean 28S/18S ratio was 1.0 (range, 0.3–1.9), and the RIN was greater than or equal to 5.0.

The sample preparation was performed according to the instructions and recommendations provided by the manufacturer. Per RNA sample, 300ng was used as input into the Affymetrix procedures as recommended by protocol (<http://www.affymetrix.com/>). RNA samples were converted to double-strand cDNA. Using a random hexamer incorporating a T7 promoter, amplified RNA (aRNA) was generated from the double-strand cDNA template through an *in vitro* transcription reaction and purified with the Affymetrix sample cleanup module. The cDNA was regenerated through a random-primed reverse transcription using dNTP mix containing dUTP. The cDNA was then fragmented by UDG and APE 1 restriction endonucleases and end-labeled by terminal transferase reaction incorporating a biotinylated dideoxynucleotide. Fragmented and end-labeled cDNAs were hybridized using the GeneChip Human Gene 2.0 ST oligonucleotide arrays (53,617 probes) for 16 hours at 45°C and 60 rpm, as described in the Gene Chip Whole Transcript (WT) Sense Target Labeling Assay Manual (Affymetrix). After hybridization, chips were stained and washed in the Genechip Fluidics Station 450 (Affymetrix). An Affymetrix Model 3000 G7 scanner and Affymetrix Command Console Software 1.1 were used for scanning and data extraction. The raw CEL file containing intensity data was used for further analysis. For normalization, the robust multiarray-average algorithm was used⁴³, which was developed in the Speed Lab at UC Berkeley.

Statistical analysis and bioinformatics analysis. Demographic characteristics, clinical, pathologic, and mammographic findings were compared between groups using the chi-square test and Fisher's exact test for categorical variables. A one-way analysis of variance (ANOVA) was performed for numerical variables.

Statistical analyses of microarray data were performed using R software, version 3.2.4 (<http://www.r-project.org/>). The R oligo package was used for processing microarray data, which is freely available on the internet (<http://www.bioconductor.org/>)^{44,45}. To identify differentially expressed genes between the three groups, a one-way ANOVA was performed with post hoc comparisons with Tukey's honest significant difference test⁴⁶, and a p value < 0.05 was applied as the threshold for statistical significance in the subsequent data analysis.

To gain insight into the underlying biology of DEGs related to mammographic calcifications, functional categories enriched in the differentially expressed genes were identified using the functional annotation and clustering tool of the Database for Annotation, Visualization, and Integrated Discovery (DAVID) v6.7 (<https://david.ncifcrf.gov/>)^{47,48}. The probability that a GO biological process term was overrepresented was determined using a modified Fisher's exact test, comparing the proportion of genes in the entire genome that are part of that GO term to the proportion of differentially expressed genes that are part of the same GO term⁴⁹.

In addition, to interpret and organize observed biological changes, we used BioLattice (<http://www.snubi.org/software/biolattice/>), a mathematical framework based on a concept lattice analysis to make associations of gene expression clusters with biological ontologies or biological pathways. BioLattice considers gene expression clusters as objects and annotations as attributes and provides a graphical summary of the order of relationships by arranging them on a concept lattice in an order based on the set inclusion relationship. Rather than interpreting one cluster at a time, BioLattice integrates all gene expression clusters and annotations into a unified framework as a lattice of concepts⁵⁰. We used Pearson correlation as a similarity measure and set arbitrary k as 60, 17 and 16 for the comparisons between breast cancers with highly suspicious calcifications and those without suspicious calcifications, breast cancers with highly suspicious calcifications and those with low-to-intermediate suspicious calcifications, and breast cancers with low-to-intermediate suspicious calcifications and those without suspicious calcifications, respectively to a cluster containing genes from 15 to 30⁵¹. We selected a threshold of $p < 0.001$. In addition, we compared our results with commercially available gene signatures used to provide recurrence score: PAM50 assay (NanoString Technologies Inc., Seattle, WA, USA)^{52,53}, MammaPrint[®] 70-gene Breast Cancer Recurrence Assay (Agendia, Huntington Beach, CA, USA)^{54,55} and The 21-gene Recurrence Score[®] assay (Oncotype DX[®], Genomic Health, Inc., Redwood City, CA, USA)^{56,57}.

Pathologic validation. Hematoxylin and eosin-stained slides of frozen human tumor tissue were examined per standard protocols for the pathologic diagnosis. Immunohistochemical analysis was performed on formalin-fixed, paraffin-embedded 4-mm tissue sections using primary mouse monoclonal antibodies for ER, PR, and HER2. For equivocal HER2 results (2+), the status was determined using fluorescence *in situ* hybridization⁵⁸.

After gene ontology and BioLattice analyses, we found that breast cancers with suspicious calcifications are associated with decreased immune system activity. Thus, we additionally planned pathologic review regarding TIL as a part of validation. One pathologist (H.S.R., with 10 years of experience) retrospectively reviewed the H&E slides of the patients and assessed TIL score using methodological recommendation of International TILs Working Group 2014⁵⁹. Of total 168 patients, pathologic slides were available for review only in 130 patients (77.4%); 78 patients (78.8%) of breast cancers without suspicious calcifications, 26 patients (70.3%) of breast cancers with low-to-intermediate suspicious calcifications, and 26 patients (81.3%) of breast cancers with highly suspicious calcifications. The mean TIL score was compared between three groups using ANOVA and independent sample t-test.

References

- Nystrom, L. *et al.* Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* **359**, 909–919, doi:10.1016/S0140-6736(02)08020-0 (2002).
- Weigel, S. *et al.* Calcifications in digital mammographic screening: improvement of early detection of invasive breast cancers? *Radiology* **255**, 738–745, doi:10.1148/radiol.10091173 (2010).
- Holland, R., Hendriks, J. H., Vebeek, A. L., Mravunac, M. & Schuurmans Stekhoven, J. H. Extent, distribution, and mammographic/histological correlations of breast ductal carcinoma *in situ*. *Lancet* **335**, 519–522 (1990).
- Cho, N. *et al.* Breast Cancer Screening With Mammography Plus Ultrasonography or Magnetic Resonance Imaging in Women 50 Years or Younger at Diagnosis and Treated With Breast Conservation Therapy. *JAMA oncology*, doi:10.1001/jamaoncol.2017.1256 (2017).
- Wang, X. *et al.* Primary breast carcinoma: association of mammographic calcifications with osteopontin expression. *Radiology* **254**, 69–78, doi:10.1148/radiol.2541090675 (2010).
- Sickles, E., D'Orsi, C. J., Bassett, L. W. *et al.* ACR BI-RADS[®] Mammography. In: *ACR BI-RADS[®] Atlas, Breast Imaging Reporting and Data System* (American College of Radiology, 2013).
- Bent, C. K., Bassett, L. W., D'Orsi, C. J. & Sayre, J. W. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. *AJR. American journal of roentgenology* **194**, 1378–1383, doi:10.2214/AJR.09.3423 (2010).
- Tabar, L. *et al.* A novel method for prediction of long-term outcome of women with T1a, T1b, and 10–14 mm invasive breast cancers: a prospective study. *Lancet* **355**, 429–433 (2000).
- Tabar, L. *et al.* Mammographic tumor features can predict long-term outcomes reliably in women with 1–14-mm invasive breast carcinoma. *Cancer* **101**, 1745–1759, doi:10.1002/cncr.20582 (2004).
- Holmberg, L. *et al.* Mammography casting-type calcification and risk of local recurrence in DCIS: analyses from a randomised study. *Br J Cancer* **108**, 812–819, doi:10.1038/bjc.2013.26 (2013).
- Ling, H. *et al.* Malignant calcification is an important unfavorable prognostic factor in primary invasive breast cancer. *Asia Pac J Clin Oncol* **9**, 139–145, doi:10.1111/j.1743-7563.2012.01572.x (2013).
- Bennett, R. L. *et al.* Pathological and mammographic prognostic factors for screen detected cancers in a multi-centre randomised, controlled trial of mammographic screening in women from age 40 to 48 years. *Breast* **20**, 525–528, doi:10.1016/j.breast.2011.05.008 (2011).

13. Gajdos, C. *et al.* Mammographic appearance of nonpalpable breast cancer reflects pathologic characteristics. *Ann Surg* **235**, 246–251 (2002).
14. Palka, I., Ormandi, K., Gaal, S., Boda, K. & Kahan, Z. Casting-type calcifications on the mammogram suggest a higher probability of early relapse and death among high-risk breast cancer patients. *Acta oncologica* **46**, 1178–1183, doi:10.1080/02841860701373611 (2007).
15. Bae, M. S. *et al.* Mammographic features of calcifications in DCIS: correlation with oestrogen receptor and human epidermal growth factor receptor 2 status. *European radiology* **23**, 2072–2078, doi:10.1007/s00330-013-2827-9 (2013).
16. Seo, B. K. *et al.* Correlation of HER-2/neu overexpression with mammography and age distribution in primary breast carcinomas. *Acad Radiol* **13**, 1211–1218, doi:10.1016/j.acra.2006.06.015 (2006).
17. Sun, S. S., Zhang, B., Zhao, H. M. & Cao, X. C. Association between mammographic features and clinicopathological characteristics in invasive ductal carcinoma of breast cancer. *Mol Clin Oncol* **2**, 623–629, doi:10.3892/mco.2014.297 (2014).
18. Wang, Y. *et al.* Estrogen receptor-negative invasive breast cancer: imaging features of tumors with and without human epidermal growth factor receptor type 2 overexpression. *Radiology* **246**, 367–375, doi:10.1148/radiol.2462070169 (2008).
19. Wang, X. *et al.* Correlation of mammographic calcifications with Her-2/neu overexpression in primary breast carcinomas. *J Digit Imaging* **21**, 170–176, doi:10.1007/s10278-008-9105-4 (2008).
20. Song, S. E. *et al.* MR and mammographic imaging features of HER2-positive breast cancers according to hormone receptor status: a retrospective comparative study. *Acta radiologica* **58**, 792–799, doi:10.1177/0284185116673119 (2017).
21. Yamamoto, S., Maki, D. D., Korn, R. L. & Kuo, M. D. Radiogenomic analysis of breast cancer using MRI: a preliminary study to define the landscape. *AJR. American journal of roentgenology* **199**, 654–663, doi:10.2214/AJR.11.7824 (2012).
22. Wan, T. *et al.* A Radio-genomics Approach for Identifying High Risk Estrogen Receptor-positive Breast Cancers on DCE-MRI: Preliminary Results in Predicting OncotypeDX Risk Scores. *Sci Rep* **6**, 21394, doi:10.1038/srep21394 (2016).
23. Zhu, Y. *et al.* Deciphering Genomic Underpinnings of Quantitative MRI-based Radiomic Phenotypes of Invasive Breast Carcinoma. *Sci Rep* **5**, 17787, doi:10.1038/srep17787 (2015).
24. Yamamoto, S. *et al.* Breast Cancer: Radiogenomic Biomarker Reveals Associations among Dynamic Contrast-enhanced MR Imaging, Long Noncoding RNA, and Metastasis. *Radiology* **275**, 384–392, doi:10.1148/radiol.15142698 (2015).
25. Evans, A. J. *et al.* Correlations between the mammographic features of ductal carcinoma *in situ* (DCIS) and C-erbB-2 oncogene expression. Nottingham Breast Team. *Clinical radiology* **49**, 559–562 (1994).
26. Rubin, I. & Yarden, Y. The basic biology of HER2. *Annals of oncology: official journal of the European Society for Medical Oncology* **12**(Suppl 1), S3–8 (2001).
27. Yepes, M. M. *et al.* Can mammographic and sonographic imaging features predict the Oncotype DX recurrence score in T1 and T2, hormone receptor positive, HER2 negative and axillary lymph node negative breast cancers? *Breast Cancer Res Treat* **148**, 117–123, doi:10.1007/s10549-014-3143-z (2014).
28. Chae, E. Y. *et al.* Association between Ultrasound Features and the 21-Gene Recurrence Score Assays in Patients with Oestrogen Receptor-Positive, HER2-Negative, Invasive Breast Cancer. *PLoS One* **11**, e0158461, doi:10.1371/journal.pone.0158461 (2016).
29. Hafez, A. *et al.* Coll11a1 Regulates Bone Microarchitecture during Embryonic Development. *Journal of developmental biology* **3**, 158–176, doi:10.3390/jdb3040158 (2015).
30. Cheon, D. J. *et al.* A collagen-remodeling gene signature regulated by TGF-beta signaling is associated with metastasis and poor survival in serous ovarian cancer. *Clinical cancer research: an official journal of the American Association for Cancer Research* **20**, 711–723, doi:10.1158/1078-0432.CCR-13-1256 (2014).
31. Vazquez-Villa, F. *et al.* COL11A1/(pro)collagen 11A1 expression is a remarkable biomarker of human invasive carcinoma-associated stromal cells and carcinoma progression. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* **36**, 2213–2222, doi:10.1007/s13277-015-3295-4 (2015).
32. Rozenchan, P. B. *et al.* Reciprocal changes in gene expression profiles of cocultured breast epithelial cells and primary fibroblasts. *International journal of cancer* **125**, 2767–2777, doi:10.1002/ijc.24646 (2009).
33. Le Bret, S. C., Newgreen, D. F., Thompson, E. W. & Ackland, M. L. Induction of epithelial to mesenchymal transition in PMC42-LA human breast carcinoma cells by carcinoma-associated fibroblast secreted factors. *Breast cancer research: BCR* **9**, R19, doi:10.1186/bcr1656 (2007).
34. Halsted, K. C. *et al.* Collagen alpha1 (XI) in normal and malignant breast tissue. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology, Inc* **21**, 1246–1254, doi:10.1038/modpathol.2008.129 (2008).
35. Freire, J. *et al.* Collagen, type XI, alpha 1: an accurate marker for differential diagnosis of breast carcinoma invasiveness in core needle biopsies. *Pathology, research and practice* **210**, 879–884, doi:10.1016/j.prp.2014.07.012 (2014).
36. Fuentes-Martinez, N. *et al.* Overexpression of proCOL11A1 as a stromal marker of breast cancer. *Histology and histopathology* **30**, 87–93, doi:10.14670/HH-30.87 (2015).
37. Tse, G. M., Tan, P. H., Cheung, H. S., Chu, W. C. & Lam, W. W. Intermediate to highly suspicious calcification in breast lesions: a radio-pathologic correlation. *Breast Cancer Res Treat* **110**, 1–7, doi:10.1007/s10549-007-9695-4 (2008).
38. Gajewski, T. F., Schreiber, H. & Fu, Y. X. Innate and adaptive immune cells in the tumor microenvironment. *Nature immunology* **14**, 1014–1022, doi:10.1038/ni.2703 (2013).
39. Disis, M. L. Immune regulation of cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **28**, 4531–4538, doi:10.1200/JCO.2009.27.2146 (2010).
40. Holland, R. & Hendriks, J. H. Microcalcifications associated with ductal carcinoma *in situ*: mammographic-pathologic correlation. *Seminars in diagnostic pathology* **11**, 181–192 (1994).
41. Stomper, P. C. & Connolly, J. L. Ductal carcinoma *in situ* of the breast: correlation between mammographic calcification and tumor subtype. *AJR. American journal of roentgenology* **159**, 483–485, doi:10.2214/ajr.159.3.1323923 (1992).
42. Imbeaud, S. *et al.* Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces. *Nucleic Acids Res* **33**, e56, doi:10.1093/nar/gni054 (2005).
43. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264, doi:10.1093/biostatistics/4.2.249 (2003).
44. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367, doi:10.1093/bioinformatics/btq431 (2010).
45. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315, doi:10.1093/bioinformatics/btg405 (2004).
46. Kerr, M. K., Martin, M. & Churchill, G. A. Analysis of variance for gene expression microarray data. *J Comput Biol* **7**, 819–837, doi:10.1089/10665270050514954 (2000).
47. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, doi:10.1038/nprot.2008.211 (2009).
48. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13, doi:10.1093/nar/gkn923 (2009).
49. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29, doi:10.1038/75556 (2000).

50. Kim, J., Chung, H. J., Jung, Y., Kim, K. K. & Kim, J. H. BioLattice: a framework for the biological interpretation of microarray gene expression data using concept lattice analysis. *J Biomed Inform* **41**, 232–241, doi:10.1016/j.jbi.2007.10.003 (2008).
51. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nat Genet* **22**, 281–285, doi:10.1038/10343 (1999).
52. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **27**, 1160–1167, doi:10.1200/JCO.2008.18.1370 (2009).
53. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752, doi:10.1038/35021093 (2000).
54. van 't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536, doi:10.1038/415530a (2002).
55. van't Veer, L. J., Paik, S. & Hayes, D. F. Gene expression profiling of breast cancer: a new tumor marker. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **23**, 1631–1635, doi:10.1200/JCO.2005.12.005 (2005).
56. Paik, S. *et al.* A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *The New England journal of medicine* **351**, 2817–2826, doi:10.1056/NEJMoa041588 (2004).
57. Paik, S. *et al.* Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **24**, 3726–3734, doi:10.1200/JCO.2005.04.7985 (2006).
58. Wolff, A. C. *et al.* American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *Arch Pathol Lab Med* **131**, 18–43, doi:10.1043/1543-2165(2007)131[18:ASOCCO]2.0.CO;2 (2007).
59. Salgado, R. *et al.* The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Annals of oncology: official journal of the European Society for Medical Oncology* **26**, 259–271, doi:10.1093/annonc/mdu450 (2015).

Acknowledgements

Sung Ui Shin is currently receiving a grant (#2015R1C1A1A01053362) from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning. For the remaining authors none were declared.

Author Contributions

W.K.M. and S.U.S. performed study design, statistical analysis, drafted the manuscript and prepared Figure 2. J.H.K. and J.L. contributed in research approaches, statistical analysis, results analysis and prepared Figures 3–4. W.H. and W.H.K. contributed in data collection. S.E.S., A.C. and W.K.M. performed mammography interpretation. H.S.R. performed pathologic slide review regarding tumor infiltrating lymphocyte. All authors contributed in editing/revising the manuscript for critical intellectual content and approved submission of the final manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-11331-9

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017