Breast Cancer: Radiogenomic Biomarker Reveals Associations among Dynamic Contrast-enhanced MR Imaging, Long Noncoding RNA, and Metastasis[1]

Shota Yamamoto, MD
Wonshik Han, MD, PhD
Youngwoo Kim, PhD
Liutao Du, MD, PhD
Neema Jamshidi, MD, PhD
Danshan Huang, MD, PhD
Jong Hyo Kim, PhD
Michael D. Kuo, MD

Purpose: To perform a radiogenomic analysis of women with breast cancer to study the multiscale relationships among quantitative computer vision–extracted dynamic contrast material–enhanced (DCE) magnetic resonance (MR) imaging phenotypes, early metastasis, and long noncoding RNA (lncRNA) expression determined by means of high-resolution next-generation RNA sequencing.

Materials and Methods: In this institutional review board–approved study, an automated image analysis platform extracted 47 computational quantitative features from DCE MR imaging data in a training set (n = 19) to screen for MR imaging biomarkers indicative of poor metastasis-free survival (MFS). The lncRNA molecular landscape of the candidate feature was defined by using an RNA sequencing–specific negative binomial distribution differential expression analysis. Then, this radiogenomic biomarker was applied prospectively to a validation set (n = 42) to allow prediction of MFS and lncRNA expression by using quantitative polymerase chain reaction analysis.

Results: The quantitative MR imaging feature, enhancing rim fraction score, was predictive of MFS in the training set (P = .007). RNA sequencing analysis yielded an average of 55.7×10^6 reads per sample and identified 14 880 lncRNAs from a background of 189 883 transcripts per sample. Radiogenomic analysis allowed identification of three previously uncharacterized and five named lncRNAs significantly associated with high enhancing rim fraction, including Homeobox transcript antisense intergenic RNA (HOTAIR) ($P < .05$), a known predictor of poor MFS in patients with breast cancer. Independent validation confirmed the association of the enhancing rim fraction phenotype with both MFS (P = .002) and expression of four of the top five differentially expressed lncRNAs ($P < .05$), including HOTAIR.

Conclusion: The enhancing rim fraction score, a quantitative DCE MR imaging lncRNA radiogenomic biomarker, is associated with early metastasis and expression of the known predictor of metastatic progression, HOTAIR.

Online supplemental material is available for this article.

[1] From the Department of Radiological Sciences, UCLA School of Medicine, Box 951721, CHS 17-135, Los Angeles, CA 90095-1721 (S.Y., L.D., N.J., D.H., M.D.K.); and Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea (W.H., Y.K., J.H.K.). Received November 21, 2014; revision requested December 8; revision received January 5, 2015; accepted January 12; final version accepted January 16. **Address correspondence to** M.D.K. (e-mail: *michaelkuo@mednet.ucla.edu*).

Long noncoding RNAs (lncRNAs), defined as noncoding transcripts greater than 200 nucleotides, represent an important class of regulatory RNAs that possess exquisite cell and tissue specificity and are critical to maintaining tissue structure and organization (1–3). Their role in breast cancer and in other solid tumors is increasingly being recognized (4–6). Indeed, dysregulated expression of the lncRNA, homeobox transcript antisense intergenic RNA (HOTAIR), has been shown to directly contribute to large-scale alterations in homeotic regulation and to be a powerful driver of metastasis in patients with breast cancer (4). Unfortunately, despite an increasing appreciation of the critical role lncRNAs play in cell fate, little is known about their relationship with systems-level phenotypes captured at noninvasive imaging.

A number of recent developments now afford an opportunity to begin exploring multiscale relationships of human breast cancers. First, advances in dynamic contrast material–enhanced (DCE) breast magnetic resonance (MR) imaging enable highly spatially and temporally resolved, intact, systems-level structural and functional characterization of breast tumors (7). Second, developments in next-generation sequencing technologies are allowing RNA sequencing at increasing resolution, yielding insights into completely new classes of transcriptional and epigenetic regulators in patients with breast cancer (8,9). Third, gains in computational power and advances in algorithm design now allow powerful computer-aided lesion detection and feature quantification and analysis of the massive datasets generated from high-throughput sequencing experiments (10,11). Finally, the emergence of radiogenomics now allows for intelligent analysis of clinical imaging data and seamless integration with clinical, tissue, cellular, and molecular phenotypes to facilitate objective, multiscale interrogation of intact systems (12–18). We performed a radiogenomic analysis of women with breast cancer to study the multiscale relationships among quantitative computer vision–extracted DCE MR imaging phenotypes, early metastasis, and lncRNA expression determined by means of high-resolution next-generation RNA sequencing.

Advances in Knowledge

- A quantitative dynamic contrast-enhanced MR imaging biomarker called the enhancing rim fraction score was associated with shorter metastasis-free survival (log rank, $P = .002$) in women with invasive ductal breast carcinoma.
- The long noncoding RNA molecular landscape of the enhancing rim fraction score was validated and shown to be significantly associated with expression of four long noncoding RNAs, including Homeobox Transcript Antisense Intergenic RNA, a powerful prognostic marker of metastasis-free survival.

Implication for Patient Care

- A quantitative dynamic contrast-enhanced MR imaging long noncoding RNA radiogenomic biomarker called the enhancing rim fraction score is associated with early metastasis and expression of the long noncoding RNA, Homeobox Transcript Antisense Intergenic RNA.

Materials and Methods

Patients and Specimens

A general overview of the analysis protocol can be seen in Figure 1. In this retrospective study, the institutional review board at Seoul National University Hospital (Seoul, Korea) approved all protocols for tissue, genomic, and clinical analysis, and written informed consent was obtained from all patients. An initial training set of data from 70 patients was procured from the Seoul National University Hospital database between January 1, 2008 and December 31, 2008. All patients had been proven to have invasive ductal carcinoma stage IIIC or lower at pathologic examination and had undergone baseline DCE MR imaging within 21 days of surgery. From this initial set of 70 patients, 19 (27.1%) were selected on the basis of meeting the following specifications: *(a)* greater than 1 μg of high-quality total RNA determined by using the standard 260/280 ratio of absorbance greater than 2.0 per sample, *(b)* RNA integrity number greater than 5, and *(c)* more than 3 months of clinical follow-up. A similar independent validation set of 75 women was selected between January 1, 2009 and July 1, 2010, 42 (56%) of whom had been proven to have invasive ductal carcinoma stage IIIC or lower at pathologic examination and had undergone baseline DCE MR imaging within 21 days of surgery; of these, 39 patients had sufficient high-quality RNA for quantitative polymerase chain reaction analysis.

Pathologic Review

Hematoxylin and eosin–stained slides of frozen human tumor tissue were examined per standard protocol for pathologic diagnosis. Immunohistochemical analysis was used to measure estrogen receptor, progesterone receptor, tumor protein 53, and epidermal growth factor receptor 2 levels. For

Published online before print
10.1148/radiol.15142698 **Content code:** BR

Radiology 2015; 275:384–392

Abbreviations:
DCE = dynamic contrast material enhanced
ERF = enhancing rim fraction
HOTAIR = Homeobox Transcript Antisense Intergenic RNA
lncRNA = long noncoding RNA
MFS = metastasis-free survival

Author contributions:
Guarantors of integrity of entire study, S.Y., M.D.K.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, S.Y., Y.K., J.H.K., M.D.K.; clinical studies, S.Y., W.H., Y.K., M.D.K.; experimental studies, S.Y., W.H., Y.K., L.D., D.H., J.H.K., M.D.K.; statistical analysis, S.Y., Y.K., L.D., N.J., J.H.K., M.D.K.; and manuscript editing, S.Y., Y.K., L.D., N.J., J.H.K., M.D.K.

Conflicts of interest are listed at the end of this article.

Figure 1

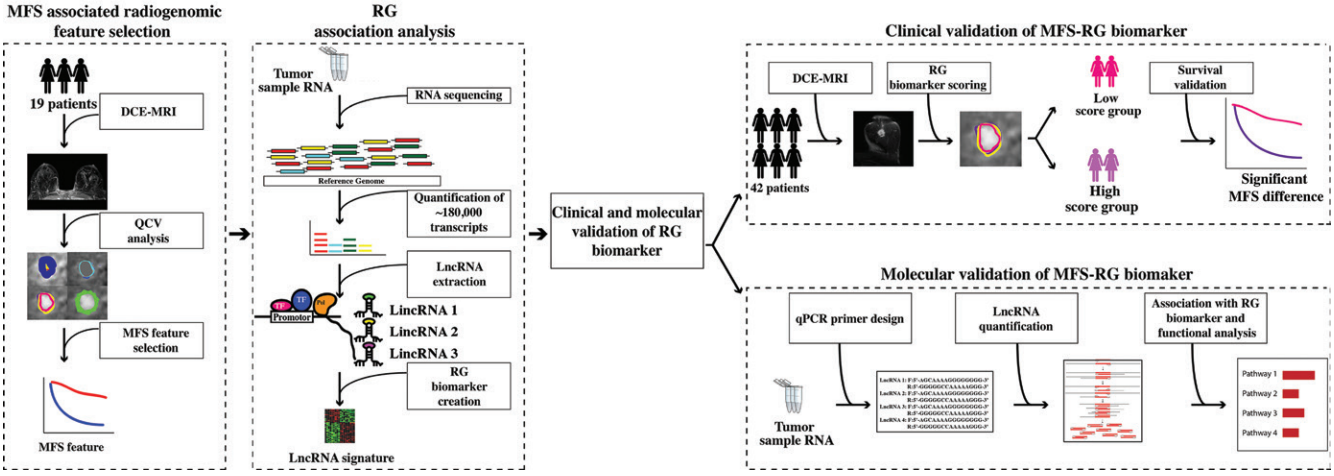

Figure 1: Illustration shows general overview of radiogenomic analysis to identify and validate a DCE MR imaging biomarker and its relationship to MFS and lncRNA expression in breast cancer. *QCV* = quantitative computer vision, *qPCR* = quantitative polymerase chain reaction, *RG* = radiogenomic.

equivocal epidermal growth factor receptor 2 results, the status was determined by means of fluorescence in-situ hybridization.

Next-Generation RNA Sequencing Protocol

Libraries were constructed with a standard library preparation kit (TruSeq RNA Illumina; San Diego, Calif) according to the manufacturer's instructions. Sequencing was performed on the largest tumor by using a sequencing platform (HiSeq 2000; Illumina) to obtain 100 base pair paired-end reads at an average of 50 million reads per patient sample. RNA sequencing reads were mapped to the human genome by using software (TopHat version 1.3.3; Center for Bioinformatics and Computational Biology, College Park, Md). A reference genome sequence (hg19, Genome Reference Consortium GRCh37) and annotation data were downloaded from the University of California, Santa Cruz Web site (*http://genome.ucsc.edu*). RNA sequencing produced an average ± standard deviation of 55.7 × 10^6 ± 4.13 reads per breast tumor. Raw counts at the transcript level (189883 elements) were extracted by using software (easyRNASeq version 2.0.8; Fred Hutchinson Cancer Research Center, Seattle, Wash) (19); 14880 lncRNAs from a genes and gene variants library (GENCODE version 7; Wellcome Trust Sanger Institute, Hinxton, England) were manually curated for further differential expression analysis. Only transcripts with nonzero raw values throughout more than 85% of the samples were included.

Real-time Quantitative Polymerase Chain Reaction Analysis

Total RNA from tumor tissues was extracted by using an RNA purification kit (RNeasy kit; Qiagen, Valencia, Calif). Polymerase chain reaction primers were designed, synthesized, and then validated in a breast MCF10a cell line (ATCC CRL-10317; American Type Culture Collection, Manassas, Va). We used 300 ng of total RNA from each tumor sample to make complementary DNA, and quantitative polymerase chain reaction was performed in triplicate in a 384-well format. The geometric mean values of three reference genes (*ACTB, HUWE1,* and *TFRC)* were used for normalization. Relative quantification values were calculated by using the δ threshold cycle method (20).

Imaging Performance and Analysis

All images were acquired by using a 1.5-T unit (Signa HDxt; GE Medical Systems, Milwaukee, Wis) with the following parameters for the bilateral protocol: repetition time msec/echo time msec, 6.2–6.5/2.2–2.5; section thickness, 3.0 mm; field of view, 18–20 cm; matrix, 320 × 224 or 320 × 150. For the unilateral protocol, the following parameters were used: 4.9/1.8; section thickness, 1.0–1.3 mm; field of view, 17–19 cm; matrix, 448 × 224. Gadopentetate dimeglumine (Magnevist; Schering, Berlin, Germany) or gadobutrol (Gadovist; Schering) was administered intravenously by means of power injection at a dose of 0.1 mmol per kilogram of body weight at a flow rate of 2 mL/sec for 5 seconds. Finally, T1-weighted three-dimensional fast low-angle shot dynamic sequences were performed, with one precontrast and four postcontrast-enhanced sagittal volume imaging series with intervals of 1.5, 3.5, 4.5, and 6.5 minutes or 1.4, 2.8, 4.2, and 5.6 minutes.

A breast DCE MR imaging segmentation and analysis software tool developed by author J.H.K.(ImagePrism4D; Seoul National University College of Medicine, Seoul, Korea) was used for imaging analysis (21). After initial automated tumor segmentation, 47 features were automatically evaluated in the tumor masks (22). The feature set consisted of three categories: geometric, statistical, and spatiotemporal

features. Two time points were selected for postcontrast volume imaging: the imaging time nearest to 2 minutes after contrast material injection was chosen as time point 1, and that nearest to 6 minutes was chosen as time point 2. If more than one lesion was identified, the largest tumor was used for extraction of features. All segmentations were independently reviewed by two expert radiologists (M.K. and N.C., with 10 and 15 years of experience, respectively). The overall automated imaging workflow is presented in Figure 2, *A*.

Geometric, Statistical, and Spatiotemporal Features

Geometric features included largest tumor volume and roundness of the largest tumor, which were computed by using volumetric segmentation data on a voxel-by-voxel basis. Statistical features were computed from the precontrast and first and second postcontrast imaging data. The features related to the intensity probability, such as standard deviation, entropy, skewness, and kurtosis, were computed by using the voxel histogram in the three-dimensional tumor (23). The features reflecting local gray-level co-occurrence matrix contrast, gray-level co-occurrence matrix homogeneity, and gray-level co-occurrence matrix energy and global (the seven Hu invariant moments) spatial patterns were extracted in the largest

Figure 2

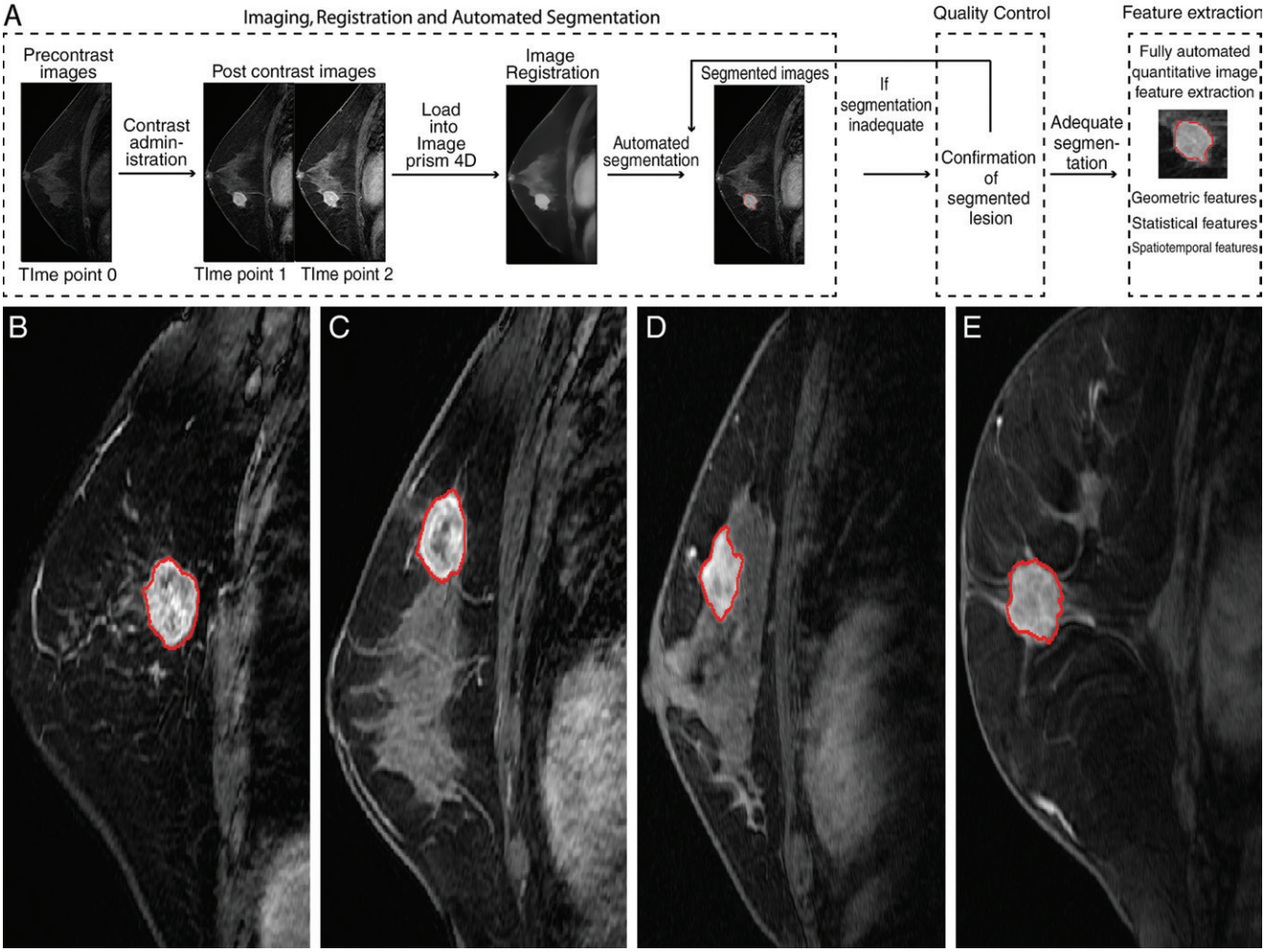

Figure 2: Sagittal images show, *A*, automated computer-vision DCE MR imaging analysis process. *B*, Image in a 67-year-old woman with estrogen and progesterone receptor–positive and human epithelial receptor type 2–negative invasive ductal carcinoma with ERF score of 0.72 (high ERF score). *C*, Image in a 56-year-old woman with estrogen and progesterone receptor–negative and human epithelial receptor type 2–negative invasive ductal carcinoma with ERF score of 0.83 (high ERF score). *D*, Image in a 64-year-old woman with estrogen and progesterone receptor–negative and human epithelial receptor type 2–positive invasive ductal carcinoma with ERF score of 0.0049 (low ERF score). *E*, Image in a 56-year-old woman with estrogen and progesterone receptor–positive and human epithelial receptor type 2 negative invasive ductal carcinoma with ERF score of 0.038 (low ERF score).

section in the three-dimensional tumor (24). Contrast enhancement-based spatiotemporal association features were calculated on the basis of classification of time-intensity curves according to the American College of Radiology Breast Imaging Reporting and Data System MR imaging guidelines and three time-point methods (25,26). Time-intensity curves were quantified for their plateau, persistent, and wash-out patterns on the basis of the time points. Additional traits such as enhancing rim fraction (ERF) and heterogeneity were measured (10,21,23). Further details of each imaging trait are provided in Table E1 (online).

The entire workflow was implemented twice by two different investigators with combined experience of 25 years to evaluate reproducibility of the image feature measurements, segmentation, and subsequent image feature analysis (J.K. and Y.K., with 20 and 5 years of experience, respectively). Co-correlation analysis on a trait-by-trait basis between the two analyses was then performed to evaluate consistency of the trait measurements. Also, to assess the repeatability of the segmentation, the spatial correlation between different statistical segmentations was analyzed by computing the Jaccard-Tanimoto coefficient. Finally, to identify the most informative traits for radiogenomic analysis, we filtered traits on the basis of their frequency, variance and prominence in the data, and independence from other traits (Pearson correlation, $r < 0.7$), and the remaining features that passed this correlation filter were used for subsequent integrative analysis. Additional details can be found in Appendix E1 (online).

Statistical Analysis

For the evaluation of metastasis-free survival (MFS), Cox regression analysis was used to identify image features that were correlated with metastasis. The resultant significant image features were then used to divide the patients in the training set into two groups on the basis of the median feature score. Kaplan-Meier plots were created, and statistical differences were evaluated by using the log-rank Mantel-Cox test. Differences between the characteristics of the training and validation set were evaluated by using the Student *t* test. Identification of differentially abundant lncRNA transcripts in the training set tumor samples was accomplished by using software (DESeq R package version 1.6.1; *www.bioconductor.org*) (27). DESeq calculates *P* values by using a negative binomial distribution, which accounts for technical and biologic variability. Differential expression was reported as fold change along with Bonferroni corrected *P* values. For our analysis, we defined a significant difference in the DESeq analysis as a *P* value less than .05 and $\log_2$ fold change of greater than 1.5. For quantitative polymerase chain reaction validation, normalized threshold cycle values were obtained, and the Spearman rank correlation coefficient was calculated, with the lncRNAs that showed a significant difference selected. Hypergeometric distribution was calculated to assess statistical significance. The lncRNAs that were successfully validated by means of quantitative polymerase chain reaction analysis were used as inputs for the lncRNA functional analysis tool (lncRNAtor version 1.0; Ewha Research Center for Systems Biology, Seoul, Korea). This tool allows functional investigation of lncRNAs by using gene expression data from 208 RNA sequencing studies (4995 samples) collected from genome databases (Gene Expression Omnibus, Encyclopedia of DNA Elements, and the Cancer Genome Atlas)(28). Software (Ingenuity Pathway Analysis; Qiagen, Redwood City, Calif) was used for lncRNA pathway enrichment analysis. Statistical analyses were performed with software (R version 3.0.1; *http://www.R-project.org* and SPSS version 21.0; IBM, Armonk, NY).

Results

Patient Characteristics

No statistically significant differences were found between the training and validation sets with respect to patient age, tumor size, stage, MFS, receptor status (estrogen receptor, progesterone receptor, human epithelial receptor type 2, tumor protein 53, triple negative status), chemotherapy, and number of tumor foci ($P > .05$). The median follow-up duration for the training and validation sets was 5.10 and 5.17 years, respectively. Detailed patient characteristics for the training and validation sets are summarized in Table 1.

Imaging Feature Analysis

The imaging features measured were highly consistent throughout the two independent analyses, with 46 of 47 (97.9%) imaging features showing significant reproducibility (Pearson correlation coefficient, $r > 0.9$; $P < .001$); the feature of tumor roundness showed a slightly lower but still significant correlation ($r = 0.78$, $P = .003$) (Table E2 [online]). The Jaccard index for the statistical traits between the two independent analyses was calculated; the mean ± standard deviation for the 39 traits was 0.80 ± 0.15. Twenty-one imaging features (two geometric, 13 statistical, and six spatiotemporal, Table E1 [online]) were used for subsequent radiogenomic analysis by applying the filtering step (Pearson correlation, $r > 0.7$), thereby eliminating closely related traits.

The ERF Phenotype Is Associated with MFS

Cox regression analysis performed on the training set by using the 21 imaging features compared with MFS allowed identification of the ERF score as the only significant imaging feature associated with early occurrence of metastasis ($P = .017$; hazard ratio, 6.06; 95% confidence interval: 1.39, 26.5). The binary cutoff value for dichotomizing patients into low- versus high-ERF score groups was 0.4 (Fig 2, *B–E*). Application of this ERF score cutoff in the training set revealed two groups (12 patients in the low ERF score group, mean ERF score of 0.25; seven patients in the high ERF score group, mean ERF score of 0.49), with an undefined median MFS time for the low ERF score (more than 50% were still alive at the time of analysis), compared with 1.8 years for the high ERF

Table 1

Demographics of Training and Validation Set

Variable	Training	Validation	*P* Value*
No. of patients	19	42	
Follow-up time (y)†	5.1 (0.83–9.4)	5.17 (0.35–8.2)	.61
Age (y)†	44 (35–70)	45 (28–75)	.86
Stage			.91
IA or IB	1	8	
IIA or IIB	12	14	
IIIA, IIIB, or IIIC	6	20	
Tumor volume (cm^3)	19.3	14.7	.65
Estrogen receptor positive	6	19	.32
Progesterone receptor positive	7	21	.35
Tumor protein 53 positive	13	24	.62
Human epithelial receptor type 2 positive	3	8	.6
Triple negative	5	10	.95
Distal recurrence event	8	18	.96
No. of patients treated with chemotherapy	17 (89)	36 (86)	.34
No. of tumor foci†	1 (1–6)	1 (1–13)	.42

Note.—Unless otherwise indicated, data are number of patients, with percentage in parentheses.

* Comparison of means by using the Student *t* test.

† Data are medians, with the range in parentheses.

Figure 3

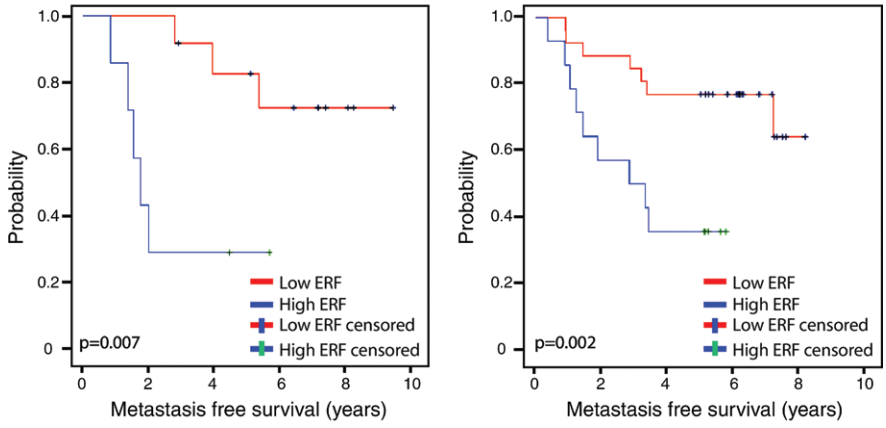

Figure 3: Graphs show MFS of training and validation sets. **(a)** Median MFS time in training set for low ERF score was undefined and for high ERF score group was 1.8 years ($P = .007$, log-rank test). **(b)** Median MFS time in validation set for low ERF score was undefined and for high ERF score group was 1.9 years ($P = .002$, log-rank test).

score group ($P = .007$, log-rank test) (Fig 3a).

Differential Expression Analysis of ERF Score Reveals the LncRNA Landscape of the ERF Score Radiophenotype

Next-generation RNA sequencing produced a mean of 189883 transcripts per sample. From this, 14880 lncRNAs were extracted and filtered down to 4468 elements per sample for further differential expression analysis. To better understand the relationships among the ERF systems-level phenotype captured at DCE MR imaging, metastasis, and lncRNA expression, a radiogenomic analysis was performed that allowed identification of eight lncRNAs at a Bonferroni-adjusted significance threshold *P* value of less than .05, and a $\log_2$ fold change of greater than 1.5. Of these, three are unnamed (signified by a code RP11 or LINC). The full list is detailed in Table 2.

Relationship between ERF Score and LncRNA HOTAIR

Because the lncRNA HOTAIR is a known regulator of metastatic progression in breast cancer, we specifically evaluated whether the ERF score was associated with HOTAIR expression. Spearman rank correlation between ERF score and HOTAIR levels demonstrated significant correlation ($r = 0.639$, $P = .003$). HOTAIR levels also were correlated with the binary ERF score ($r = 0.538$, $P = .018$).

Validation of the ERF Score with MFS

Cox regression analysis performed on the validation set by using the ERF scores compared with MFS showed significant differences in MFS ($P = .009$; hazard ratio, 16.3; 95% confidence interval: 1.99, 133.04). Application of the prespecified ERF cutoff value established in the training set resulted in 27 patients in the low– and 15 patients in the high–ERF score groups. We found that the binary ERF score remained a significant predictor of MFS ($P = .002$, log-rank test), with an undefined median survival time (greater than 50% alive at the time of analysis) versus 1.9 years for the low– versus high–ERF score groups, respectively (Fig 3b).

Validation of the Relationship between the ERF Score and LncRNA Expression

Next, we sought to confirm prospectively the radiogenomic associations of the ERF phenotype with its predicted lncRNA expression. Sufficient quality total RNA was available for analysis for 39 of 42 (92.8%) validation set patients and was used for quantitative polymerase chain reaction validation. Of the eight lncRNAs identified in the training set, we selected the top five differentially expressed lncRNAs ($\log_2$ fold change above 2.2) and constructed and validated polymerase chain reaction primers for validation in the independent set (Table E3 [online]). Prospective quantitative polymerase chain

reaction analysis on the validation set samples confirmed a significant association expression of four of five of these lncRNAs (RP11–278 L15.2–001, LINC00511–009, HOTAIR, AC004231.2–001; $P < .05$) with the ERF score in the independent validation set (hypergeometric distribution: 80% observed vs < 1% expected by chance alone; $P = 1.76 \times 10^{-4}$), thus confirming the association of the ERF score with its predicted lncRNA expression correlates. We again observed that the ERF score was directly correlated with lncRNA HOTAIR expression ($P = .025$), and binary ERF score ($P = .047$) in the independent validation set, confirming the relationship between this powerful regulator of metastasis in patients with breast cancer and the ERF score. The overall molecular landscape of the ERF score is presented in Figure 4.

Functional Investigation of the Validated LncRNAs

Finally, in an effort to characterize our validated lncRNAs associated with the ERF phenotype, we applied two different bioinformatics approaches. First, we used software (lncRNAtor version 1.0; Ewha Research Center for Systems Biology) as a tool to survey the Cancer Genome Atlas breast cancer data for determining the implication of the four lncRNAs. Two of the validated lncRNAs, LINC00511 and HOTAIR, were found to be present in higher levels in tumor samples of invasive ductal carcinoma of the breast than in normal breast tissue on the basis of 894 samples (LINC00511 fold change, 6.25; $P < .001$; HOTAIR fold change, 5.56; $P < .001$; respectively). LINC00511 was also enriched significantly in tumor samples of kidney, head and neck, liver, lung, and bladder cancers compared with their respective normal tissue samples ($P < .001$). In addition, Ingenuity pathway analysis was performed in which four associated significant pathways were identified ($P < .05$). These pathways included cell cycle, cell death and survival, cellular development, and cellular growth and proliferation.

Table 2

Differential Expression of lncRNA, Low versus High ERF

Symbol	Ensemble Transcript ID	Log_2 Fold Change	Adjusted *P* Value*	Gene Biotype
RP11–278 L15.2–001	ENST00000462931	4.446	<.001	long noncoding
LINC00511–009	ENST00000457958	2.728	<.001	long noncoding
HOTAIR	ENST00000424518	2.258	<.001	antisense
AC004231.2–001	ENST00000418393	2.207	<.001	antisense
RP11–731F5.2–002	ENST00000460164	2.088	<.001	long noncoding
WWTR1-AS1–003	ENST00000495094	1.796	.017	antisense
BCYRN1–001	ENST00000418539	1.677	.010	long noncoding
ST8SIA6-AS1–001	ENST00000457649	1.633	.014	antisense

* Corrected with Benjamini-Hochberg method.

Figure 4

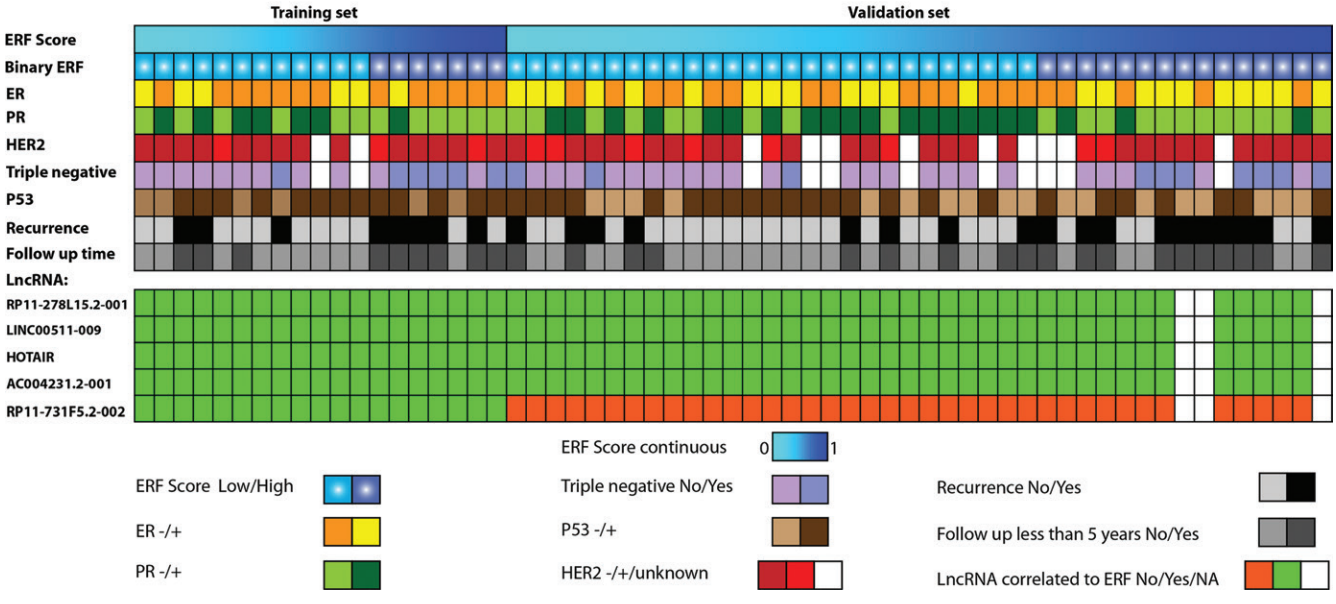

Figure 4: Graph shows molecular characteristics of the ERF score phenotype in training ($n = 19$) and validation ($n = 42$) sets. Continous ERF scores for each patient are listed from low to high in the respective data sets. Status of estrogen receptor, progesterone receptor, epidermal growth factor receptor 2, triple negative receptor, tumor protein 53, and lncRNA expression are provided as labeled. Recurrence and follow-up data are also included.

Discussion

In this radiogenomic study of breast cancer, we integrated DCE MR imaging with lncRNA profiling to identify radiogenomic biomarkers of early breast cancer metastasis by using automated quantitative feature extraction, next-generation tumor RNA sequencing, and quantitative polymerase chain reaction validation. A feature called the ERF score was identified, and patients with high ERF scores were found to experience metastasis earlier than patients with low ERF scores. In addition, RNA sequencing analysis of 4468 lncRNAs allowed identification of eight lncRNAs that were strongly associated with the ERF phenotype. These included several unknown lncRNAs and known lncRNAs including HOTAIR, a known driver of metastasis (4). The ERF score radiogenomic associations were validated in an independent dataset, confirming their association with MFS and with expression of lncRNAs by means of quantitative polymerase chain reaction. Thus, these findings begin to reveal an initial portrait of the multiscale relationships between quantitative DCE MR imaging phenotypes, early metastasis, and lncRNA expression observed at high resolution RNA sequencing, which suggests that the approach is both feasible and scalable.

As with other imaging modalities, there is a need for automated computer–vision and quantitative feature–extraction algorithms in breast DCE MR imaging. Although authors of previous work, to our knowledge, have relied on detailed image object identification followed by expert-guided segmentation and feature analysis, the analysis system used in our study is almost completely automated (29,30). This increases objectivity, thereby rendering spatially and temporally dynamic noninvasive image phenotyping more akin to molecular profiling tools. Furthermore, we used a relatively focused, yet diverse image feature library that included statistical, morphologic, and textural features as well as spatiotemporally based image features (31–33). The statistical and textural measures were found to be uninformative, but the spatiotemporal-type trait, ERF, was both strongly associated with MFS and biologically concordant with known lncRNA regulators of aggressive tumors. Although further work is currently underway to ensure reproducibility of our approach, these initial results are promising.

We identified the ERF phenotype to be significantly associated with the lncRNA HOTAIR, which has been implicated in breast cancer metastasis (4). Data suggest that HOTAIR functions by transducing higher-order spatial information in cells by mediating epigenetic silencing in trans through physical recruitment of polycomb repressive complex 2 (PCR2) and subsequent targeting of tumor suppressor genes (34). Thus, HOTAIR appears to effect cellular reprogramming by reimposing the chromatin state to one associated with dedifferentiated, migratory, and invasive fibroblasts, consistent with its association with metastasis (4). Furthermore, HOTAIR also is known to increase expression of several angiogenic genes such as vascular endothelial growth factor (35). Given the contrast-enhanced phenotypic characteristics of the high-ERF trait and its strong association with early metastasis found in our study, these findings appear to be compatible with known effects of HOTAIR. Clearly, further work is necessary to better understand the potential mechanistic relationships between HOTAIR expression and the ERF phenotype.

Despite our efforts to minimize biases in our study through the use of computer-aided lesion identification and quantitative feature extraction, a uniform patient population, standardized imaging protocols, and independent training and validation sets, care must be taken not to overgeneralize our results. Limitations include the retrospective, single-institution nature of the study. Further validation in a larger cohort of prospectively recruited patients of heterogeneous ethnic backgrounds from different institutions is required and will ensure robustness. Further studies also will be required for confirmation of any mechanistic relationships between the associations of imaging features with the identified lncRNAs. Although future radiogenomic studies of breast cancer could benefit from the leveraging of existing large public datasets such as the Cancer Genome Atlas and the Cancer Imaging Archive, they currently are limited by nonstandardized image acquisition protocols, small sample size, and immature outcomes data (8). Nonetheless, we expect that, with time, such resources will become an invaluable resource for future radiogenomic analyses of breast cancer.

The growing number of genomic, imaging, and clinical biomarkers identified in patients with breast cancer has created the need for integrative biomarkers, biomarkers that link multiple types of data and measurements. Although it is beyond the scope of this analysis, this radiophenotypic map of breast cancer can serve as a fundamental starting point with which to build even more in-depth hierarchical biologic models that reflect how the ERF phenotype integrates with the network of other molecular species such as mRNA and microRNA expression, DNA methylation, and DNA copy number and sequence variation. These newly identified radiogenomic associations set the stage for future studies to elucidate the roles and functions of new, previously uncharacterized lncRNAs and other genomic elements and their relationship to systems-level image phenotypes in a multiscale context.

Disclosures of Conflicts of Interest: S.Y. disclosed no relevant relationships. **W.H.** disclosed no relevant relationships. **Y.K.** disclosed no relevant relationships. **L.D.** disclosed no relevant relationships. **N.J.** disclosed no relevant relationships. **D.H.** disclosed no relevant relationships. **J.H.K.** disclosed no relevant relationships. **M.D.K.** disclosed no relevant relationships.

References

1. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem 2012;81:145–166.
2. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009;10(3):155–159.
3. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet 2011;12(12):861–874.

4. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature 2010;464(7291):1071–1076.
5. Spizzo R, Almeida MI, Colombatti A, Calin GA. Long non-coding RNAs and cancer: a new frontier of translational research? Oncogene 2012;31(43):4577–4587.
6. Brunner AL, Beck AH, Edris B, et al. Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. Genome Biol 2012;13(8):R75.
7. O'Connor JP, Jackson A, Parker GJ, Jayson GC. DCE-MRI biomarkers in the clinical evaluation of antiangiogenic and vascular disrupting agents. Br J Cancer 2007;96(2):189–195.
8. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490(7418):61–70.
9. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet 2010;11(1):31–46.
10. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioglu O, Su MY. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. Acad Radiol 2008;15(12):1513–1525.
11. Chen W, Giger ML, Newstead GM, et al. Computerized assessment of breast lesion malignancy using DCE-MRI robustness study on two independent clinical datasets from two manufacturers. Acad Radiol 2010;17(7):822–829.
12. Diehn M, Nardini C, Wang DS, et al. Identification of noninvasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A 2008;105(13):5213–5218.
13. Jamshidi N, Diehn M, Bredel M, Kuo MD. Illuminating radiogenomic characteristics of glioblastoma multiforme through integration of MR imaging, messenger RNA expression, and DNA copy number variation. Radiology 2014;270(1):1–2.
14. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic analysis to identify imaging phenotypes associated with drug response gene expression programs in hepatocellular carcinoma. J Vasc Interv Radiol 2007;18(7):821–831.
15. Kuo MD, Jamshidi N. Behind the numbers: Decoding molecular phenotypes with radiogenomics—guiding principles and technical considerations. Radiology 2014;270(2):320–325.
16. Yamamoto S, Korn RL, Oklu R, et al. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. Radiology 2014;272(2):568–576.
17. Yamamoto S, Maki DD, Korn RL, Kuo MD. Radiogenomic analysis of breast cancer using MRI: a preliminary study to define the landscape. AJR Am J Roentgenol 2012;199(3):654–663.
18. Segal E, Sirlin CB, Ooi C, et al. Decoding global gene expression programs in liver cancer by noninvasive imaging. Nat Biotechnol 2007;25(6):675–680.
19. Delhomme N, Padioleau I, Furlong EE, Steinmetz LM. easyRNASeq: a bioconductor package for processing RNA-Seq data. Bioinformatics 2012;28(19):2532–2533.
20. Nygard AB, Jørgensen CB, Cirera S, Fredholm M. Selection of reference genes for gene expression studies in pig tissues using SYBR green qPCR. BMC Mol Biol 2007;8(1):67.
21. Lee SH, Kim JH, Cho N, et al. Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. Med Phys 2010;37(8):3940–3956.
22. Hong BW. Joint estimation of shape and deformation for the detection of lesions in dynamic contrast-enhanced breast MRI. Phys Med Biol 2013;58(21):7757–7775.
23. Fitzpatrick JM, Sonka M. Handbook of Medical Imaging, Vol 2: Medical Image Processing and Analysis. Bellingham, WA: SPIE Press, 2000.
24. Hu MK. Visual pattern recognition by moment invariants. IRE Trans Inf Theory 1962;8(2):179–187.
25. Degani H, Gusis V, Weinstein D, Fields S, Strano S. Mapping pathophysiological features of breast tumors by MRI at high spatial resolution. Nat Med 1997;3(7):780–782.
26. Furman-Haran E, Feinberg MS, Badikhi D, Eyal E, Zehavi T, Degani H. Standardization of radiological evaluation of dynamic contrast enhanced MRI: application in breast cancer diagnosis. Technol Cancer Res Treat 2014;13(5):445–454.
27. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11(10):R106.
28. Park C, Yu N, Choi I, Kim W, Lee S. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. Bioinformatics 2014;30(17):2480–2485.
29. Yankeelov TE, Lepage M, Chakravarthy A, et al. Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results. Magn Reson Imaging 2007;25(1):1–13.
30. Park SH, Moon WK, Cho N, et al. Diffusion-weighted MR imaging: pretreatment prediction of response to neoadjuvant chemotherapy in patients with breast cancer. Radiology 2010;257(1):56–63.
31. Woods BJ, Clymer BD, Kurc T, et al. Malignant-lesion segmentation using 4D co-occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data. J Magn Reson Imaging 2007;25(3):495–501.
32. Holli K, Lääperi AL, Harrison L, et al. Characterization of breast cancer types by texture analysis of magnetic resonance images. Acad Radiol 2010;17(2):135–141.
33. Agner SC, Rosen MA, Englander S, et al. Computerized image analysis for identifying triple-negative breast cancers and differentiating them from other molecular subtypes of breast cancer on dynamic contrast-enhanced MR images: a feasibility study. Radiology 2014;272(1):91–99.
34. Tsai MC, Manor O, Wan Y, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science 2010;329(5992):689–693.
35. Geng YJ, Xie SL, Li Q, Ma J, Wang GY. Large intervening non-coding RNA HOTAIR is associated with hepatocellular carcinoma progression. J Int Med Res 2011;39(6):2119–2128.