

# Radiogenomic Analysis of Breast Cancer by Linking MRI Phenotypes with Tumor Gene Expression

Tycho Bismeyer, PhD • Bas H. M. van der Velden, PhD • Sander Canisius, PhD • Esther H. Lips, PhD • Claudette E. Loo, MD, PhD • Max A. Viergever, PhD • Jelle Wesseling, MD, PhD • Kenneth G. A. Gilhuijs, PhD • Lodewyk F. A. Wessels, PhD

From the Division of Molecular Carcinogenesis, Oncode Institute (T.B., S.C., L.F.A.W.), Division of Molecular Pathology (S.C., E.H.L., J.W.), Department of Radiology (C.E.L.), and Department of Pathology (J.W.), the Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands; Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands (B.H.M.v.d.V., M.A.V., K.G.A.G.); and Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Delft, the Netherlands (L.F.A.W.). Received July 8, 2019; revision requested September 11; final revision received March 13, 2020; accepted March 27. **Address correspondence to** L.F.A.W. (e-mail: [L.wessels@nki.nl](mailto:L.wessels@nki.nl)).

This research is part of the STW Perspectief Population Imaging Genetics (ImaGene) program and supported by the Dutch Technology Foundation STW, which is part of the Netherlands Organisation for Scientific Research (NWO), and partly funded by the Ministry of Economic Affairs.

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

See also the editorial by Cho in this issue.

Radiology 2020; 00:1–11 • <https://doi.org/10.1148/radiol.2020191453> • Content codes: **BR** **MR**

**Background:** Better understanding of the molecular biology associated with MRI phenotypes may aid in the diagnosis and treatment of breast cancer.

**Purpose:** To discover the associations between MRI phenotypes of breast cancer and their underlying molecular biology derived from gene expression data.

**Materials and Methods:** This is a secondary analysis of the Multimodality Analysis and Radiologic Guidance in Breast-Conserving Therapy, or MARGINS, study. MARGINS included patients eligible for breast-conserving therapy between November 2000 and December 2008 for preoperative breast MRI. Tumor RNA was collected for sequencing from surgical specimen. Twenty-one computer-generated MRI features of tumors were condensed into seven MRI factors related to tumor size, shape, initial enhancement, late enhancement, smoothness of enhancement, sharpness, and sharpness variation. These factors were associated with gene expression levels from RNA sequencing by using gene set enrichment analysis. Statistical significance of these associations was evaluated by using a sample permutation test and the false discovery rate.

**Results:** Gene expression and MRI data were obtained for 295 patients (mean age, 56 years  $\pm$  10.3 [standard deviation]). Larger and more irregular tumors showed increased expression of cell cycle and DNA damage checkpoint genes (false discovery rate  $<0.25$ ; normalized enrichment statistic [NES], 2.15). Enhancement and sharpness of the tumor margin were associated with expression of ribosomal proteins (false discovery rate  $<0.25$ ; NES, 1.95). Smoothness of enhancement, tumor size, and tumor shape were associated with expression of genes involved in the extracellular matrix (false discovery rate  $<0.25$ ; NES, 2.25).

**Conclusion:** Breast cancer MRI phenotypes were related to their underlying molecular biology revealed by using RNA sequencing. The association between enhancements and sharpness of the tumor margin with the ribosome suggests that these MRI features may be imaging biomarkers for drugs targeting the ribosome.

© RSNA, 2020

Online supplemental material is available for this article.

Dynamic contrast material-enhanced (DCE) MRI is used for the screening and diagnosis of breast cancer (1). Using radiomics, images from MRI examinations are condensed into a set of numerical features that describe the phenotype of the tumor. These features aim to characterize morphologic findings, enhancement, enhancement dynamics, and texture of tumors (2–5). More MRI features are published every year, but the biologic mechanisms that explain the clinical association of these features remain unexplored.

MRI features describe tumor characteristics that are the result of biologic processes in the tumor and hence can aid in investigating the mechanisms that drive tumor progression or therapy resistance. For example, tumor size, which can be measured with MRI, is the result of biologic processes such as cell proliferation and cell death. Therefore,

MRI features that show chance association with clinical outcome and do not validate in an independent data set are less likely to be associated with biologic processes. An MRI feature with a biologic interpretation is much more likely to validate successfully in an independent patient cohort and can be prioritized for validation.

Carcinogenesis occurs at the molecular level. It attains the hallmarks of cancer (6) by generating genetic and epigenetic aberrations. For genetic and epigenetic aberrations to be carcinogenic, they need to result in aberrant protein function. These are mediated through, or eventually produce, aberrant gene expression (RNA expression levels). Gene expression can be measured in tumor tissue obtained from biopsies or surgery specimens. It has been used for its diagnostic and prognostic value (7,8). Gene expression also reflects the biologic processes active in a tumor, which can

## Abbreviations

DCE = dynamic contrast enhanced, NES = normalized enrichment statistic

## Summary

MRI-derived radiomics features were associated with protein expression related to the ribosome (targeted in anticancer drugs) and genes related to the extracellular matrix (involved in cancer progression).

## Key Results

- The MRI phenotypes low initial enhancement, increased smoothness of enhancement, and low sharpness are associated with the expression of proteins that are part of the ribosome (normalized enrichment statistic [NES], 1.95), a target of anticancer drugs under active investigation.
- Increased smoothness of enhancement, smaller tumor size, and a more irregular tumor shape are associated with the expression of genes related to the extracellular matrix (NES, 2.25), which is involved in breast cancer progression and metastasis.

be elucidated with a pathway analysis (9). This is a hypothesis-generating approach (10).

We hypothesize that MRI phenotypes of breast cancer can be given a biologic explanation by integrating gene expression with DCE MRI. Therefore, by performing a pathway analysis on these data, we aim to discover the associations between MRI phenotypes of breast cancer and their underlying molecular biology derived from gene expression data.

## Materials and Methods

We acquired matched breast DCE MRI and tumor gene expression data for 295 patients. DCE MRI and gene expression were jointly analyzed, and we used a pathway analysis to identify pathways that are associated with MRI features. These pathway associations allow the formulation of hypotheses regarding the biology underlying MRI features, which can help in their translation to clinical practice.

## Cohort

After institutional review board approval and written informed consent, 598 patients with breast cancer were consecutively included as part of the Multimodality Analysis and Radiologic Guidance in Breast-Conserving Therapy, or MARGINS, study between 2000 and 2008 at the Netherlands Cancer Institute (Amsterdam, the Netherlands). Patients in the MARGINS study were eligible for breast-conserving therapy based on conventional imaging and clinical assessment and were recruited for additional preoperative breast MRI (11). Hence, MRI data were available for all these patients. Image-guided fine-needle aspiration or core biopsy was used to confirm the presence of breast cancer. Treatment plans were established by a multidisciplinary team of breast cancer specialists. Neoadjuvant systemic therapy was an exclusion criterion, because this is not compatible with analysis of treatment-naïve tumor tissue. For 384 patients (64%), breast DCE MRI could be matched to tumor material from surgical excision available for research purposes in the Netherlands Cancer Institute biobank. Tumor material was retrospectively acquired with institutional review board ap-

proval. Written consent was not required under Dutch regulations (opt-out). For 303 patients (51%), the tumor material yielded enough high-quality RNA for sequencing. Among these patients, for 295 all MRI features could be computed.

We have reported on the MRI data of the MARGINS study before (Appendix E1 [online]), but these were not aimed at radiogenomic analysis.

## DCE MRI and MRI Features

The MRI protocol was described previously (12). Briefly, images were acquired by using a 1.5-T scanner (Magnetom; Siemens, Erlangen, Germany) with a dedicated double-breast array coil (CP Breast Array, four channels; Siemens) (C.L., a dedicated breast radiologist with more than 15 years of experience). A coronal fast low-angle shot three-dimensional T1-weighted series were acquired consisting of one image before contrast material injection and four images obtained after contrast material injection. Imaging parameters were as follows: acquisition time, 90 sec; repetition time, 8.1 msec; echo time, 4.0 msec; flip angle, 20°; voxel size,  $1.35 \times 1.35 \times 1.35$  mm<sup>3</sup>; field of view, 310 mm. Images from one examination were registered to each other by using deformable registration (13), after which the breast area was segmented automatically in three dimensions and the tumor was segmented in four dimensions (14).

A representative set of 21 MRI features that quantify lesion characteristics was computed by using previously published methods (3,15). These features are often used for computer-aided diagnosis in breast cancer and include features for tumor morphology and enhancement dynamics. The features were calculated with in-house software in C++ by using the equations listed in Table E1 (online).

## Factor Analysis of MRI Features

Correlation between MRI features (16) partly reflects dependencies between the features that have a purely technical origin. For example, the correlation between tumor size and almost every other feature can be explained by the statistical dependence between tumor size and the number of voxels these features are computed on. This is even true for features that seem to be invariant to tumor size because they are computed and averaged over all voxels in a tumor. The number of voxels inside the tumor grows proportionally with the cube of the tumor diameter, while the number of voxels at the edge of the tumor grows proportionally with the square of tumor diameter. Therefore, their proportion is correlated to tumor volume, which affects the calculation of tumor features. For example, if only the voxels at the edge of a tumor are affected by partial volume effects, the influence of these effects will depend on tumor size. Such statistical dependencies between MRI features do not reflect the biologic processes captured by the MRI features. Therefore, these dependencies pose a problem for integrating MRI features and data representative of the biologic processes in a tumor, such as gene expression data. For example, as tumor size is strongly associated with proliferation (shown below), the correlation between tumor size and almost every other feature leads to an association between proliferation and almost all features. These trivial associations can

obscure biologically relevant associations and need to be dealt with in a radiogenomic analysis.

To address this problem, we performed a factor analysis to remove correlation between MRI features (17) by using Python (version 3.6.0; Python Software Foundation, Wilmington, Del; open source; <https://www.python.org/>). It consists of four steps. Similar features, such as length and volume, are measured on scales that are nonlinear transformations of each other, such as millimeters and cubic millimeters. Therefore, the first step transforms all features to comparable scales by taking the square root of features that measure variance to match the scale of features that measure the mean, by taking the cubic root of features that describe volume to match features describing length, and by linearly normalizing all features to equal variance. The second step of the factor analysis is a principal component analysis of the rescaled features with the scikit-learn package (version 0.18; open source; <https://scikit-learn.org/>). We chose principal component analysis because of its simplicity and stability compared with other factor analysis methods. In the third step, we chose the number factors based on not including factors adding very little (<1%) additional variation, as the rest is likely to contain mostly noise. The fourth step of the factor analysis rotates the factors by using the varimax, quartimin, quartimax, or equamax rotations with the factor\_rotation package (open source; [https://github.com/mvds314/factor\\_rotation](https://github.com/mvds314/factor_rotation)). These rotations increase the interpretability of the factors by minimizing the number of features associated with a factor. Different rotations often give similar results, and we quantified their similarity by using the Pearson correlation of factors. All factor rotation methods result in almost equal factors (Pearson correlation >0.999;  $P < .001$ ), so we chose to continue with the results of only one of them, the varimax rotation.

### Gene Expression

Gene expression was measured by using RNA sequencing, as described previously (18). Briefly, RNA was extracted from fresh-frozen surgical excision specimen from the Netherlands Cancer Institute biobank. Samples with low tumor percentage (<30%) or low RNA quality (Bioanalyzer 2100; Agilent Technologies, Santa Clara, Calif) (RNA integrity number <6) were excluded, leaving 304 patients with material for RNA sequencing. RNA was sequenced on the HiSeq 2500 (Illumina, San Diego, Calif) with single-end 65 base-pair reads. RNA sequencing reads were aligned with STAR 2.5.0a (open source; <https://github.com/alexdobin/STAR>) to the human genome (GENCODE 23; number of genes is 60 669) and read counts per gene were determined.

### Pathway Analysis

We used pathway analyses of gene expression data to identify the biologic processes that are associated with an MRI factor. The pathway analysis starts by regressing the MRI factors on the expression of all genes. This regression determines, for a given MRI factor-gene pair, the strength of the association between a gene and a factor. This is quantified by the regression coefficient and the associated  $t$  statistic (19). The  $t$  statistics of

a single MRI factor determine a ranking of all genes, with the most strongly associated genes at the top. Based on this ranking, the pathway analysis scores the enrichment of every gene set. For example, if all genes of a given gene set are ranked at the top, then such a gene set will be scored as enriched, indicating a strong association between the gene set, or biologic process it represents, and the factor.

In addition to statistical significance, the pathway analysis results in three scores for every MRI factor-gene set pair to quantify the different aspects of the association between a factor and a gene set. First, the normalized enrichment statistic (NES) is the effect size of the gene set enrichment that is comparable between gene sets. The larger the NES, the stronger the association of a factor with a gene set. Second, the maximum enrichment statistic at measures the proportion of genes that are at least as strongly correlated with the MRI factor as the genes that contribute to the enrichment. It could be that a statistically significant enrichment is achieved, but of all the genes correlated with the factor, only a small fraction is in the gene set. In this case, the maximum enrichment statistic at is high, indicating there might be more biology associated with the factor than that described by the gene set. Third, the leading edge is the proportion of genes in a gene set that contribute to the enrichment. It could be that a statistically significant enrichment is achieved, but only a small fraction of the genes in a gene set contribute to that enrichment. In this case, the leading edge is low, indicating that the name of the gene set may be not descriptive of the enrichment.

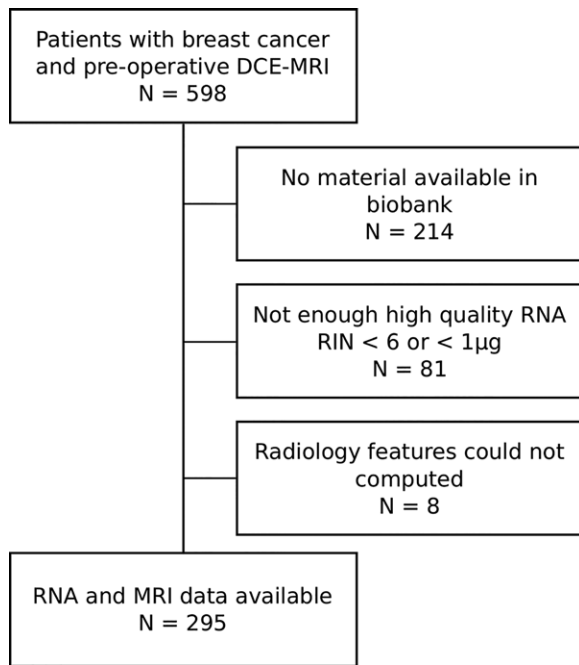
### Software Availability

All code used to generate the results in this article is available (<https://github.com/NKI-CCB/imagenome-analysis>). The flexgsea package for pathway analysis is also available (<https://github.com/NKI-CCB/flexgsea-r>).

### Statistical Analyses

In the pathway analysis, for the linear regression step, we used limma-voom (open source) (19) because standard linear regression is inappropriate for RNA-seq data (19). For the gene-set enrichment analysis (9), we used version 1.1 of the flexgsea R package (open source; <https://zenodo.org/record/1182639#.XblfePco9cU>). We used two gene set collections from the Molecular Signatures Database, version 5.2 (9): one containing experimentally derived gene sets (c2.cgp;  $n = 3400$ ) and one containing gene sets from pathway databases curated by molecular biology experts (c2.cp;  $n = 1329$ ). Combined, these two sets are comprehensive without including too many gene sets that are highly similar. The genes in the gene sets can be looked up on the Molecular Signatures Database website (<https://software.broadinstitute.org/gsea/msigdb/>). Importantly, enrichment scores were tested for significance by using a sample permutation test (false discovery rate <0.25).

All other statistical analyses were performed in Python with the scipy (version 0.19.0; open source; <https://www.scipy.org/>) and statsmodels (version 0.8.0; open source; <https://www.statsmodels.org/>) packages. Correlations were measured by using Pearson correlation and a  $P$  value cutoff of .05 was used. Significance of the



**Figure 1:** Flow diagram shows patient exclusion. DCE = dynamic contrast enhanced, RIN = RNA integrity number.

difference between tumors or patients with and without gene expression was tested with the Wilcoxon rank-sum test.

## Results

We obtained gene expression and MRI data for 295 patients, a subset of the whole cohort of 598 with DCE MRI data available (Fig 1). The clinical characteristics between patients with and those without gene expression were not significantly different (Table 1). From the DCE MRI data we calculated a representative set of 21 MRI features. The MRI features and factors between tumors with and those without gene expression show differences (Table 2; Table E2 [online]). The largest difference is in tumor size, as measured by volume, which is larger in tumors with gene expression.

### Factor Analysis

The MRI features were designed to quantify a wide range of spatial and temporal characteristics of the tumor. Hence, our data show a strong degree of correlation between many of the features (Fig 2a). Therefore, the MRI features were condensed into factors, which minimizes the correlation between them. We found that a minimum of seven factors are necessary to describe the data (Fig 2b). The size factor is a linear combination of diameter and cubic root-transformed volume features. The MRI factors are related to the MRI features by the loadings (Fig 2c; Tables 3, 4). The loadings can be used to calculate the MRI factors from the MRI features by taking the dot product between the MRI features and the loadings. The loadings measure the association between MRI factors and MRI features. The factors summarize groups of correlated MRI features, which is clearly reflected by the fact that the loadings of correlated features are typically similar within the factor summarizing them. Factor 1 is related to

**Table 1: Clinical Characteristics**

Parameter	Full Cohort	With Gene Expression
No. of patients	598	295
Mean age (y)*	56 (26–84)	58 (31–84)
Mean tumor size (mm)*	21 (5–90)	22 (6–90)
Grade		
1	195 (33)	89 (30)
2	250 (42)	135 (45)
3	148 (25)	77 (26)
Lymph node status		
Negative	389 (66)	191 (64)
1	103 (18)	49 (16)
2	42 (7)	29 (10)
3	18 (3)	8 (3)
4+	33 (6)	23 (8)
IHC subtype		
ER positive	433 (75)	238 (79)
ER negative and HER2 positive	74 (13)	30 (10)
Triple negative	69 (12)	33 (11)
Radiation therapy	512 (85)	266 (88)
Adjuvant chemotherapy	179 (30)	94 (31)
Hormonal therapy	222 (37)	128 (42)
HER2-positive targeted therapy	25 (4)	13 (4)

Note.—Unless otherwise specified, data are numbers, with percentages in parentheses. ER = estrogen receptor, HER2 = receptor tyrosine-protein kinase *ERBB2*, IHC = immunohistochemistry.

\* Data in parentheses are ranges.

tumor size as it has substantial loadings for the MRI features volume (0.45), volume where initial enhancement is larger than 100% (0.41), volume where late enhancement is smaller than 0% (0.36), largest diameter (0.39), largest diameter where initial enhancement is larger than 100% (0.38), and largest diameter where late enhancement is smaller than 0% (0.39). Factor 2 is related to initial enhancement as it has substantial loadings for the MRI features uptake speed (−0.63) and top initial enhancement (−0.77). Factor 3 is related to smoothness as it has substantial loadings for the MRI features maximum mean and variation of smoothness (−0.54, −0.47) and mean and variation of smoothness at uptake (−0.53, −0.46). Factor 4 is related to late enhancement, as it has substantial loadings for the MRI features top late enhancement (0.54), signal enhancement ratio (−0.59), and washout (−0.54). Factor 5 is related to variation of sharpness, as it has substantial loadings for the MRI features variation of sharpness at uptake (−0.69) and variation of sharpness maximum (−0.71). Factor 6 is related to shape as it has substantial loadings for the MRI features circularity (−0.79) and irregularity (0.54). Finally, factor 7 is related to mean sharpness as it has substantial loadings for the MRI features mean of sharpness at uptake (−0.68) and mean of sharpness maximum (−0.71).

### Pathway Analysis

Having established a compact representation of the MRI features in terms of MRI factors, we can turn to the main purpose of this work, namely to establish associations between MRI



**Table 2: Mean, Standard Deviation, 25% Quantiles, and 75% Quantiles of MRI Features for the Group of Patients with MRI Features Available and for the Group of Patients with Gene Expression Data Available**

MRI Feature	Full Cohort with MRI Feature ( <i>n</i> = 566)				With Gene Expression ( <i>n</i> = 295)			
	Mean	Standard Deviation	25% Quantile	75% Quantile	Mean	Standard Deviation	25% Quantile	75% Quantile
Volume (mm <sup>3</sup> )	1951	2209	642	2518	2158	2208	801	2845
Diameter (mm)	20.49	7.81	14.83	25.20	21.00	7.03	15.83	25.34
Volume initial enhancement greater than 100 (mm <sup>3</sup> )	1234	1648	304	1475	1432	1816	418	1744
Diameter initial enhancement greater than 100 (mm)	19.21	7.89	13.62	23.68	19.82	7.07	14.61	23.91
Volume late enhancement less than 0 (mm <sup>3</sup> )	784	1150	164	928	941	1294	218	1165
Diameter late enhancement less than 0 (mm)	18.56	7.81	13.07	23.07	19.12	7.12	14.04	23.52
Smoothness uptake (mean)	0.57	0.03	0.54	0.59	0.57	0.03	0.55	0.59
Smoothness uptake (variation)	3.40	1.40	2.32	4.41	3.48	1.32	2.49	4.45
Smoothness maximum (mean)	0.57	0.03	0.54	0.59	0.57	0.03	0.55	0.60
Smoothness maximum (variation)	3.42	1.42	2.33	4.44	3.62	1.36	2.61	4.57
Sharpness uptake (mean)	4.91	0.80	4.43	5.46	4.93	0.78	4.52	5.44
Sharpness uptake (variation)	6.54	2.70	4.66	7.92	6.46	2.55	4.66	7.65
Sharpness maximum (mean)	4.90	0.80	4.41	5.46	5.00	0.71	4.56	5.45
Sharpness maximum (variation)	6.58	2.64	4.80	7.81	6.55	2.48	4.83	7.66
Uptake speed	1.69	0.56	1.32	2.05	1.72	0.53	1.35	2.05
Top initial enhancement	2.38	0.77	1.85	2.81	2.37	0.72	1.84	2.82
Top late enhancement	-0.25	0.10	-0.32	-0.18	-0.25	0.10	-0.32	-0.18
Signal enhancement ratio	1.37	0.39	1.11	1.57	1.37	0.37	1.12	1.58
Washout	0.13	0.11	0.05	0.20	0.13	0.11	0.06	0.20
Circularity	0.76	0.10	0.71	0.83	0.78	0.08	0.72	0.84
Irregularity	0.44	0.08	0.38	0.48	0.43	0.07	0.38	0.49

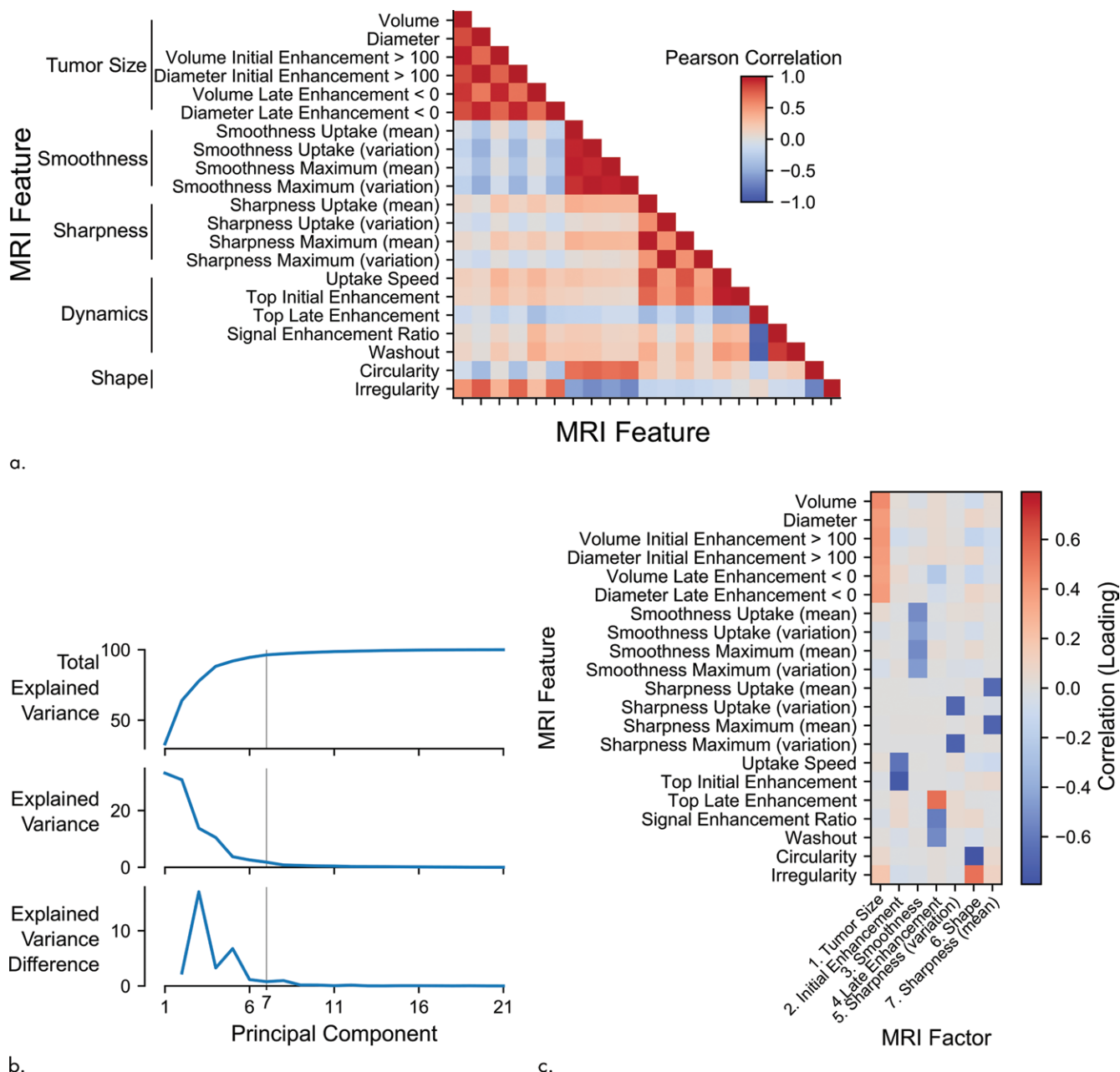
phenotypes of breast cancer and their underlying molecular biology derived from gene expression data. To this end, we use pathway analysis of the genes whose gene expression values show association with MRI features.

Pathway analysis shows that tumor size, as represented by factor 1, is associated with a gene set of doxyrubicin treatment in a retinoblastoma transcriptional corepressor 1-positive cell line (gene set I in Fig 3; NES, 1.73), a gene set of basal-like breast cancer (gene set II; NES, 1.70), a gene set of luminal A breast cancer (gene set III; NES, 1.65), two gene sets of cell cycle (gene sets IV and V; NES, 2.05 and 1.95), and a gene set of DNA damage repair genes (gene set VI; NES, 1.93) (Fig 3; false discovery rate <0.25). Based on the conditions under which these gene sets show regulation, we can conclude that they collectively represent regulation of the cell cycle and DNA damage pathways. The gene sets shown in Figure 3 have some of the strongest associations with tumor size (high NES), and are sensitive (high maximum enrichment score at) and specific (high leading edge) to tumor size (Fig 3, A) compared with other experimentally derived gene sets. The two pathways share many genes, such as those related to cell cycle checkpoints. This link between tumor size and proliferation is also seen at histopathologic analysis, as tumor size is positively correlated with tumor grade ( $P < .001$ ; Kruskal-Wallis test), which is strongly associated with proliferation (20). Importantly, the association between tumor size

and proliferative pathways cannot be explained by the estrogen receptor-positive and triple-negative subtypes. Although triple-negative tumors are highly proliferative and tend to be larger (Fig E1 [online];  $P < .01$ ; Kruskal-Wallis test) than are estrogen receptor-positive tumors, this does not explain the association, because we observe the same when analyzing only estrogen receptor-positive tumors (Figs E2, E3 [online]).

Tumors with lower initial enhancement (factor 2), low sharpness of the border (factor 3), or with smooth enhancement (factor 7) show enrichment for a gene set of the ribosome (gene set VII; NES: 1.90, 1.59, and 1.86 for factor 2, 3, and 7, respectively) and a gene set of peptide chain elongation (gene set VIII; NES: 1.86, 1.57, and 1.86), which indicates a higher expression of protein coding genes of the ribosome (Fig 4; false discovery rate <0.25). The gene sets shown in Figure 4 have some of the strongest associations with these factors, and are sensitive and specific to them, compared with other gene sets from expert-curated databases (Fig 4, A). A low sharpness of the border means that neighboring voxels at the tumor border have similar intensities (Fig 5). A high smoothness of enhancement means the enhancement seen after administration of the contrast agent has a gradient that points on average away from the tumor center (Fig 6); it is also known as the radial gradient index (3).

Finally, smaller (factor 1), less smoothly enhancing (factor 5), and more irregular and less circular (factor 6) tumors show



**Figure 2:** Images show MRI features are correlated and can be summarized in small number of factors. **(a)** Heat map of Pearson correlation between MRI features. Groups of similar features are annotated with their common property. Groups of strongly anticorrelated and correlated features are visible as blue and red blocks. **(b)** Total explained variation (proportion of MRI feature data that MRI factors capture), explained variation of single factor and difference in explained variation with preceding factor, all shown as function of number of selected factors. After inclusion of seven factors, almost no additional variance is explained (<1%), indicating that number of factors can be limited to seven without losing important information. **(c)** Heat map of loadings of MRI factors shows relationships between MRI features and MRI factors. Bright blue and red values in heat map represent large negative and positive contributions (loadings) of MRI features to MRI factors. Groups of features with similar loadings, indicating they are summarized by the same factor, correspond to correlated groups of features depicted in **a**.

enrichment for two gene sets of collagens (gene set IX; NES: 1.47, 1.97, and 2.24 for factor 1, 5, and 6, respectively and gene set X; NES: 1.56, 2.00, and 2.12) and a gene set of proteoglycan (gene set XI; NES: 1.57, 1.91, and 1.80), which indicates higher expression of genes involved in the extracellular matrix (Fig 7; false discovery rate <0.25). The gene sets shown in Figure 6 have some of the strongest associations with these factors, and are sensitive and specific to them, compared with other gene sets from expert-curated databases (Fig 7, A).

## Discussion

We used a set of previously defined dynamic contrast material-enhanced (DCE) MRI radiomics features and gene expression profiling of our patient cohort. We derived a set of factors as a compact representation of the DCE MRI features. These factors show strong association with specific gene expression patterns, and we identified the biologic processes underlying these gene expression patterns. These associations establish links between the molecular biology of three key

**Table 3: Loadings of MRI Features in MRI Factors**

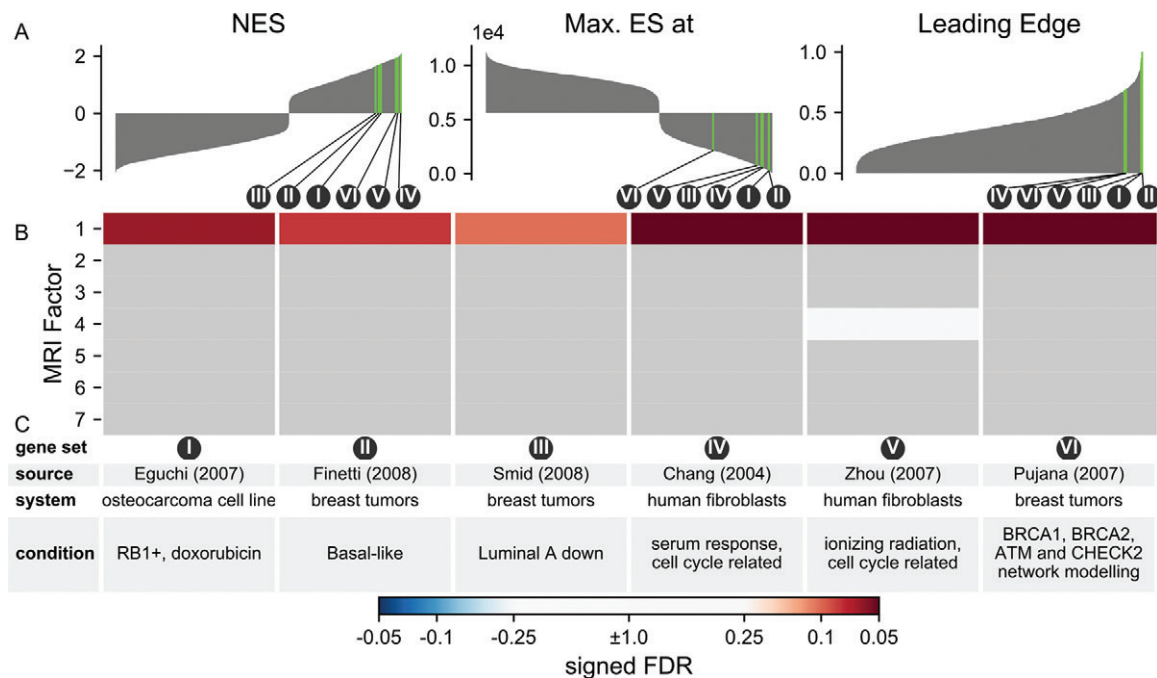
MRI Feature	Factor						
	1	2	3	4	5	6	7
Circularity	0.068	−0.015	−0.003	0.022	−0.008	−0.792	0.069
Irregularity	0.190	−0.056	−0.032	0.023	−0.021	0.535	0.107
Volume	0.450	0.030	−0.026	0.053	−0.014	−0.087	0.043
Diameter	0.391	0.012	0.034	0.055	−0.008	0.086	0.036
Uptake speed	0.026	−0.626	0.002	−0.006	0.033	−0.062	−0.094
Washout	0.016	−0.039	0.009	−0.536	−0.001	−0.048	0.009
Signal enhancement ratio	−0.033	0.063	−0.016	−0.588	0.050	0.069	−0.013
Top initial enhancement	−0.019	−0.765	−0.002	−0.000	−0.020	0.036	0.061
Top late enhancement	0.004	0.056	−0.019	0.542	0.054	−0.012	−0.016
Volume initial enhancement greater than 100%	0.412	−0.069	−0.031	0.044	0.009	−0.145	−0.081
Diameter initial enhancement greater than 100%	0.385	−0.004	0.037	0.0601	0.035	0.077	−0.057
Volume late enhancement less than 0%	0.360	0.061	−0.022	−0.235	−0.004	−0.129	−0.035
Diameter late enhancement less than 0%	0.385	0.0120	0.012	−0.058	−0.013	0.083	0.041
Sharpness maximum (mean)	−0.001	0.009	0.007	0.017	−0.001	0.028	−0.705
Sharpness maximum (variation)	−0.001	−0.006	−0.000	−0.000	−0.713	0.004	0.016
Sharpness uptake (mean)	0.006	0.004	−0.009	−0.016	−0.008	0.024	−0.676
Sharpness uptake (variation)	0.001	0.005	0.002	0.004	−0.693	−0.007	−0.026
Smoothness maximum (variation)	−0.047	0.019	−0.468	0.006	−0.027	−0.037	−0.006
Smoothness maximum (mean)	0.015	−0.005	−0.538	0.034	−0.0002	0.049	0.004
Smoothness uptake (variation)	−0.034	0.006	−0.456	−0.032	−0.005	−0.061	0.002
Smoothness uptake (mean)	0.045	−0.016	−0.527	−0.013	0.026	0.035	−0.002

Note.—See Figure 2c.

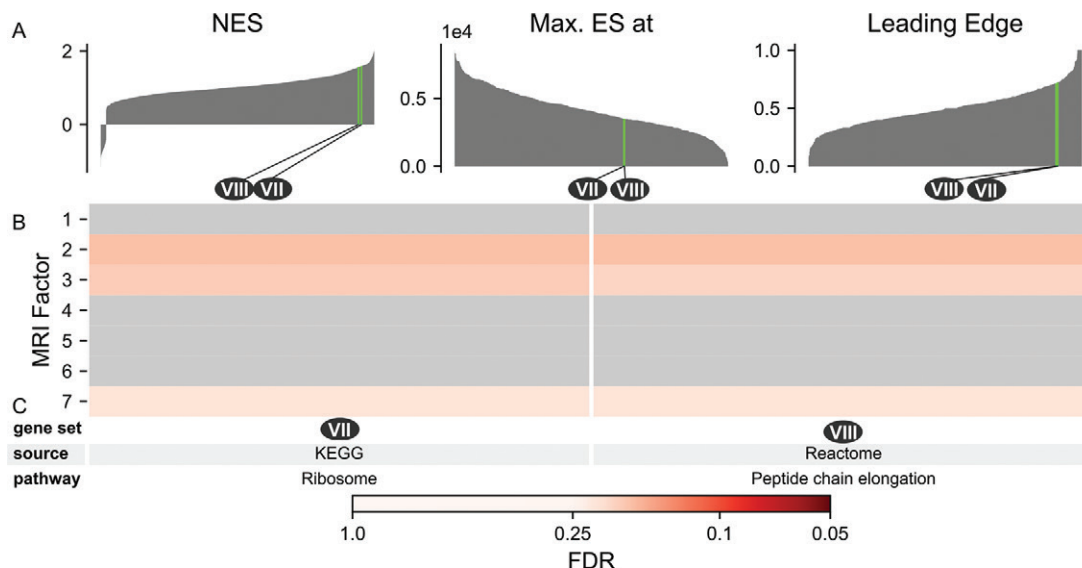
**Table 4: Loadings of MRI Features in MRI Factors Calculated on Estrogen Receptor–Positive Tumors**

MRI Feature	Factor						
	1	2	3	4	5	6	7
Circularity	0.007	−0.055	0.017	−0.017	0.010	0.804	0.044
Irregularity	−0.007	−0.193	0.007	−0.021	0.018	−0.493	0.074
Volume	−0.011	−0.458	0.044	−0.055	0.007	0.116	0.032
Diameter	0.031	−0.391	0.028	−0.057	0.014	−0.089	0.036
Uptake speed	−0.008	−0.032	−0.624	0.008	−0.028	0.046	−0.077
Washout	0.013	−0.013	−0.076	0.528	−0.006	0.062	0.048
Signal enhancement ratio	−0.009	0.032	0.105	0.579	−0.047	−0.061	−0.043
Top initial enhancement	0.004	0.021	−0.761	−0.007	0.016	−0.044	0.053
Top late enhancement	−0.013	−0.019	0.051	−0.543	−0.055	0.037	−0.005
Volume initial enhancement greater than 100%	−0.018	−0.412	−0.079	−0.041	−0.018	0.172	−0.160
Diameter initial enhancement greater than 100%	0.036	−0.383	−0.002	−0.064	−0.039	−0.079	−0.060
Volume late enhancement less than 0%	−0.027	−0.349	0.038	0.269	0.001	0.119	−0.024
Diameter late enhancement less than 0%	−0.025	−0.387	−0.011	0.055	0.031	−0.126	0.082
Sharpness maximum (mean)	0.002	0.002	0.017	−0.015	−0.001	−0.019	−0.715
Sharpness maximum (variation)	−0.002	−0.002	−0.002	−0.000	0.716	−0.000	0.016
Sharpness uptake (mean)	−0.009	−0.005	−0.006	0.021	0.011	−0.022	−0.665
Sharpness uptake (variation)	0.003	0.002	0.001	−0.003	0.691	0.009	−0.027
Smoothness maximum (variation)	−0.464	0.051	0.026	−0.012	0.018	0.033	−0.026
Smoothness maximum (mean)	−0.545	−0.016	0.012	−0.036	0.003	−0.057	−0.005
Smoothness uptake (variation)	−0.448	0.038	−0.014	0.031	−0.001	0.068	0.013
Smoothness uptake (mean)	−0.532	−0.047	−0.022	0.017	−0.017	−0.029	0.018

Note.—See Figure E3, C (online).



**Figure 3:** Images of gene set enrichment analysis of MRI factors show that tumor size (MRI factor 1) is associated with proliferation. A, Waterfall plots show statistics for MRI factor 1 (tumor size) used to select gene sets for display in heat map and table, for all gene sets in MsigDB c2.cgp, with selected gene sets in green. Waterfall plots show three types of association statistics on y-axis calculated for all gene sets, which are ordered in increasing level of association along x-axis: normalized enrichment statistic (NES), maximum enrichment statistic at (Max. ES at), and leading edge. Fill of maximum enrichment statistic waterfall plot is zero at middle, such that enrichment at top is shown downward and enrichment at bottom upwards. B, Heat map with false discovery rate (FDR) of positive (red) and negative (blue) associations between MRI factors and gene sets. Insignificant associations are shown in gray. C, Descriptions of publication, experimental system, and conditions associated with given gene set.

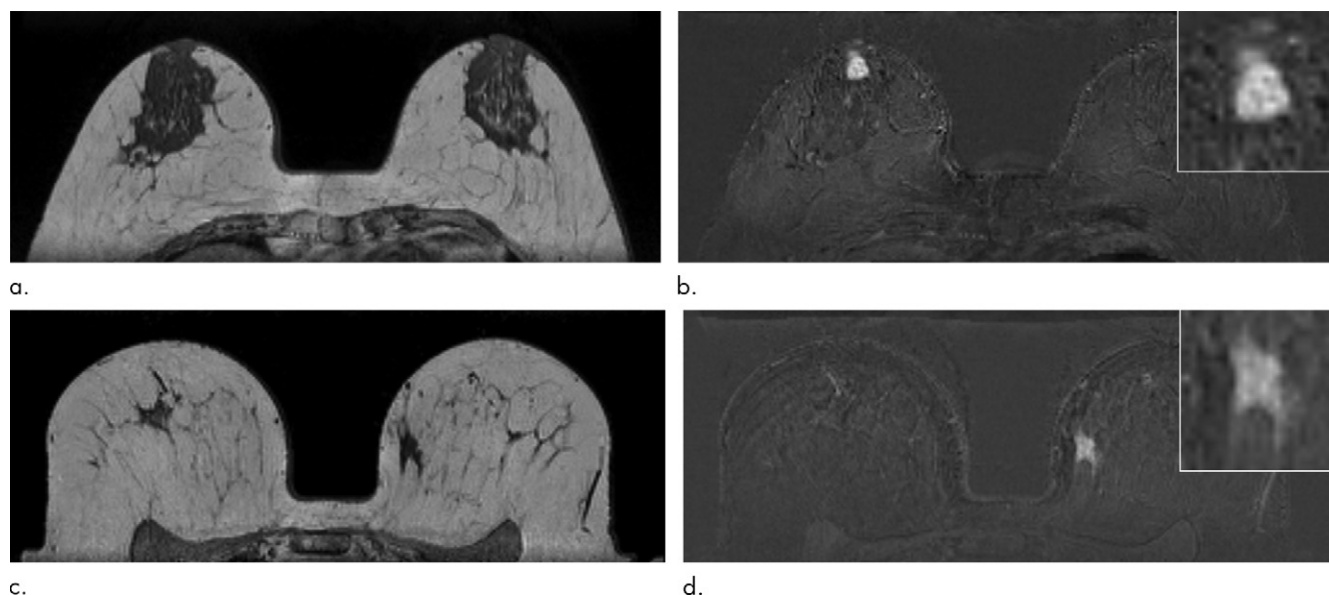


**Figure 4:** Images of gene set enrichment analysis of MRI factors show that factor 2 (initial enhancement), factor 3 (smoothness), and factor 7 (mean sharpness) are associated with ribosome. A, Waterfall plots show statistics for factor 3, used to select gene sets for display in heat map and table, for all gene sets in MsigDB c2.cp, with selected gene sets in green. Waterfall plots show three types of association statistics on y-axis calculated for all gene sets, which are ordered in increasing level of association along x-axis: normalized enrichment statistic (NES), maximum enrichment statistic at (Max. ES at), and leading edge. B, Heat map with false discovery rate (FDR) of positive (red) and negative (blue) associations between MRI factors and gene sets. Insignificant associations are shown in gray. C, Experiments that are source of gene sets.

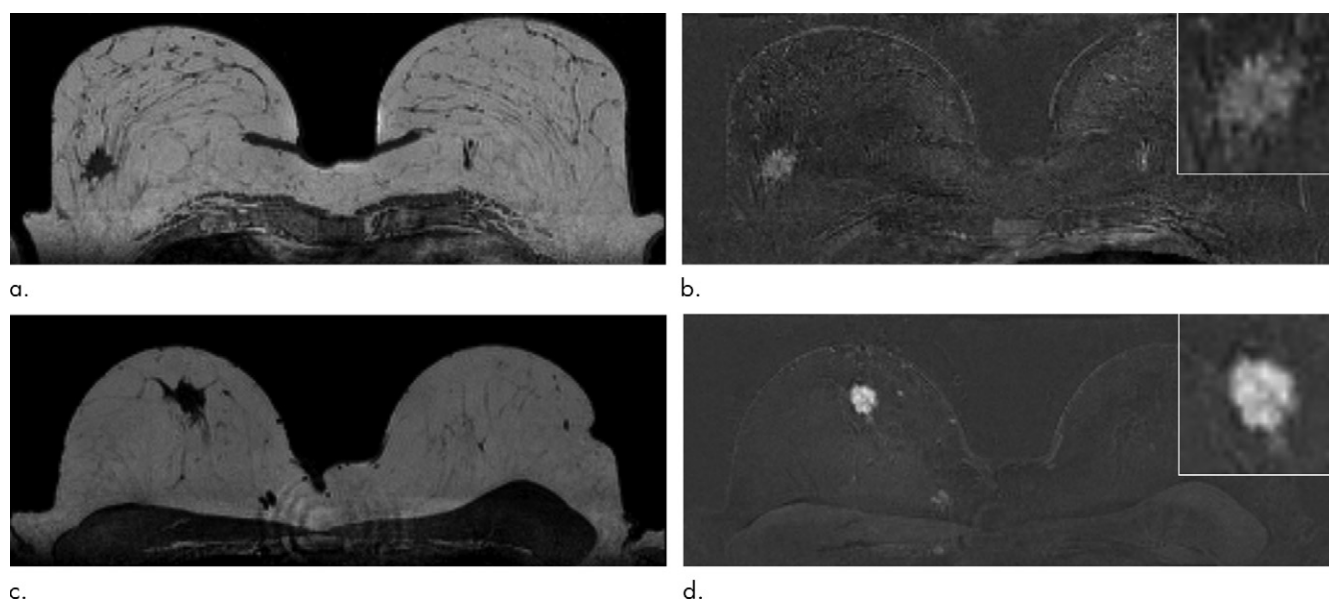
biologic processes in cancer (proliferation, protein synthesis, and the extracellular matrix) and four characteristics captured by the MRI factors (tumor size, smoothness, mean sharpness, and tumor shape).

The strongest effect we observe is the well-known association between proliferation and tumor size. This finding links two known associations. First, highly proliferative tumors are generally associated with poor outcome (21). Second,





**Figure 5:** Axial images show examples of variation in MRI factor 3 (smoothness). **(a)** MRI before contrast material injection in a 57-year-old patient with node-negative, estrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative, grade 2 invasive ductal carcinoma of 1.8 cm. Tumor has low MRI factor 3 ( $-2.35$ ), hence high smoothness. **(b)** Subtraction between first MRI after and MRI before contrast material injection in the same patient. **(c)** MRI before contrast material injection in 64-year-old patient with node-negative, ER-positive, HER2-negative, grade 2 invasive ductal carcinoma of 1.7 cm. Tumor has high MRI factor 3 ( $2.29$ ), hence low smoothness. **(d)** Subtraction between first MRI after and MRI before contrast material injection in the same patient.



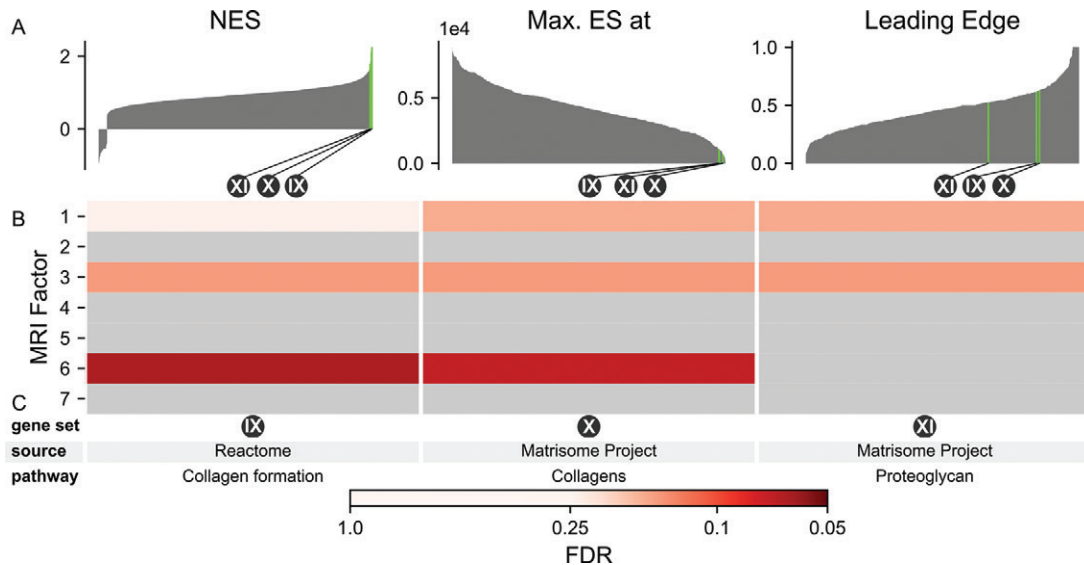
**Figure 6:** Axial images show examples of variation in MRI factor 7 (sharpness). **(a)** MRI before contrast material injection in a 71-year-old patient with estrogen receptor (ER)-positive, human epidermal growth factor receptor 2 (HER2)-negative, grade 2 invasive ductal carcinoma of 2.4 cm. Four lymph nodes tested positive for malignancy. Cancer had high MRI factor 7 ( $2.17$ ), hence low margin sharpness. **(b)** Subtraction between first MRI after and MRI before contrast material injection in the same patient. **(c)** MRI before contrast material injection in a 63-year-old patient with node-negative, ER-positive, HER2-negative, grade 1 invasive ductal carcinoma of 1.3 cm. This patient had low MRI factor 7 ( $-3.00$ ), hence high margin sharpness. **(d)** Subtraction between first MRI after and MRI before contrast material injection in the same patient.

larger tumors have a worse prognosis than do smaller tumors (22).

We found that initial enhancement, smoothness of enhancement, and sharpness of the tumor boundary were found to be associated with expression of ribosomal proteins. Ribosomal proteins are required for ribogenesis, a target for anticancer drugs (23). Furthermore, through activation of ribosomal protein S6 kinase (RPS6KB1), ribogenesis is regulated by the mTOR

pathway (24). The mTOR pathway can be targeted by drugs that have been approved or are under clinical development and is frequently deregulated in breast cancer through mutations in *PIK3CA*, *PIK3RI*, *PTEN*, and *AKT* (25).

Finally, we found that smoothness of enhancement and tumor shape are associated with expression of genes involved in the extracellular matrix and collagen production. Collagens are produced by fibroblasts, which are contributors to cancer progression (26).



**Figure 7:** Images of gene set enrichment analysis of MRI factors show that factor 3 (smoothness) and factor 6 (shape) are associated with extracellular matrix. A, Waterfall plots show statistics for factor 6, used to select gene sets for display in heat map and table, for all gene sets in MsigDB c2.cp, with selected gene sets in green. Waterfall plots show three types of association statistics on y-axis calculated for all gene sets, which are ordered in increasing level of association along x-axis: normalized enrichment statistic (NES), maximum enrichment statistic at (Max. ES at), and leading edge. B, Heat map with false discovery rate (FDR) of positive (red) and negative (blue) associations between MRI factors and gene sets. Insignificant associations are shown in gray. C, Experiments that are the source of gene sets.

By linking DCE MRI features to biologic processes, we reveal their biologic underpinnings and clinical relevance. More specifically, the link between initial enhancement, smoothness of enhancement, and sharpness of the tumor boundary and the mTOR pathway suggests that these DCE MRI-derived factors can provide a noninvasive imaging biomarker that identifies tumors susceptible to drugs that target ribosome biogenesis and the mTOR pathway.

Others have established a relationship between tumor size and prognostic gene expression signatures (27), as well as prognostic gene expression signatures and proliferation (28). The association between tumor size and proliferation we have uncovered agrees with these findings. Previously, highly enhancing parts of a mouse tumor have been reported to express RNA coding for ribosomal proteins at high levels (29). We confirm these findings by establishing a link between initial enhancement, low margin sharpness, and smoothness of enhancement with expression of ribosomal proteins. In contrast to studies that linked DCE MRI features with gene expression signatures (27,30,31), we considered the complete transcriptome and not only a subset of genes contained in the signatures. In contrast to other studies (32,33), we perform the pathway analyses with gene set enrichment analysis and subject permutation. This approach does not feature the erroneous assumption of independence between genes, but rather assumes independence between patients.

Our study also had some limitations. The choice of gene expression data over other molecular data types limited the degree to which the biology of tumors is captured. DNA mutation and copy number data would allow identification of causal drivers of DCE MRI phenotypes. Protein and phosphoprotein expression data would allow assessing signaling pathway activation. Also, gene expression was measured in a single piece of

tumor, so it does not reflect the intratumor heterogeneity that DCE MRI can capture. For only half the patients, we could acquire gene expression data, and this was biased toward patients with larger tumors, potentially limiting the applicability of the results for patients with very small tumors. A limited number of MRI features were chosen, and these were developed to support radiologists in making a diagnosis and do not capture all the information available in a DCE MRI.

Because the gene expression profile does not capture the full tumor heterogeneity, it would be of interest to relate spatially resolved molecular measurements such as topographic single cell sequencing (34) of tumors to the DCE MRI scans. In addition, more information could be extracted from the images by directly predicting the expression of cancer pathways from the images (using, for example, deep learning). We expect that such an approach could reveal many additional relevant associations. Based on our results, we speculated that deregulation of the mTOR pathway may lead to an increased smoothness of enhancement at DCE MRI. Because the activation of this pathway cannot be determined from gene expression data, we propose to validate this finding on a data set with matched breast DCE MRI and tumor DNA sequencing or tumor phosphoproteomics data.

From these three findings, we can conclude that the integration of gene expression with dynamic contrast material-enhanced (DCE) MRI has been able to generate testable hypothesis regarding possible mechanisms by which MRI features attain their clinical benefit. Therefore, the gene expression data have added substantially to the biologic understanding of DCE MRI phenotypes. Hence, searching for similar relationships in other cancer types, or with other imaging modalities, would be interesting.

**Acknowledgments:** We acknowledge Lennart Mulder, BASc, and Frank Nieboer, MSc, as well as the Core Facility Molecular Pathology and Biobank, the Genomics Core Facility, and the Research High Performance Computing facility at the Netherlands Cancer Institute.

**Author contributions:** Guarantors of integrity of entire study, T.B., E.H.L., L.F.A.W.; 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, T.B., M.A.V.; experimental studies, B.H.M.v.d.V., E.H.L., J.W.; statistical analysis, T.B., S.C., L.F.A.W.; and manuscript editing, T.B., B.H.M.v.d.V., S.C., E.H.L., C.E.L., M.A.V., J.W., L.F.A.W.

**Disclosures of Conflicts of Interest:** T.B. disclosed no relevant relationships. B.H.M.v.d.V. disclosed no relevant relationships. S.C. disclosed no relevant relationships. E.H.L. disclosed no relevant relationships. C.E.L. disclosed no relevant relationships. M.A.V. disclosed no relevant relationships. J.W. disclosed no relevant relationships. K.G.A.G. disclosed no relevant relationships. L.F.A.W. Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: has grants/grants pending with Genmab. Other relationships: disclosed no relevant relationships.

## References

- Mann RM, Cho N, Moy L. Breast MRI: State of the Art. *Radiology* 2019;292(3):520–536.
- Sinha S, Lucas-Quesada FA, DeBruhl ND, et al. Multifeature analysis of Gd-enhanced MR images of breast lesions. *J Magn Reson Imaging* 1997;7(6):1016–1026.
- Gilhuijs KGA, Giger ML, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Med Phys* 1998;25(9):1647–1654.
- Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* 2003;50(1):92–98.
- Dalmış MU, Gubern-Mérida A, Vreemann S, Karsseneijer N, Mann R, Platel B. A computer-aided diagnosis system for breast DCE-MRI at high spatiotemporal resolution. *Med Phys* 2016;43(1):84–94.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144(5):646–674.
- Parker JS, Mullins M, Cheang MCU, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* 2009;27(8):1160–1167.
- Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 2016;375(8):717–729.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102(43):15545–15550.
- Pinker K, Chin J, Melsaether AN, Morris EA, Moy L. Precision Medicine and Radiogenomics in Breast Cancer: New Approaches toward Diagnosis and Treatment. *Radiology* 2018;287(3):732–747.
- Elshof LE, Rutgers EJ, Deurloo EE, et al. A practical approach to manage additional lesions at preoperative breast MRI in patients eligible for breast conserving therapy: results. *Breast Cancer Res Treat* 2010;124(3):707–715.
- van der Velden BH, Dmitriev I, Loo CE, Pijnappel RM, Gilhuijs KG. Association between parenchymal enhancement of the contralateral breast in dynamic contrast-enhanced MR imaging and outcome of patients with unilateral invasive breast cancer. *Radiology* 2015;276(3):675–685.
- Dmitriev ID, Loo CE, Vogel WV, Pengel KE, Gilhuijs KGA. Fully automated deformable registration of breast DCE-MRI and PET/CT. *Phys Med Biol* 2013;58(4):1221–1233.
- Alderliesten T, Schlieff A, Peterse J, et al. Validation of semiautomatic measurement of the extent of breast tumors using contrast-enhanced magnetic resonance imaging. *Invest Radiol* 2007;42(1):42–49.
- Gilhuijs KGA, Deurloo EE, Muller SH, Peterse JL, Schultze Kool LJ. Breast MR imaging in women at increased lifetime risk of breast cancer: clinical system for computerized assessment of breast lesions initial results. *Radiology* 2002;225(3):907–916.
- Benndorf M, Baltzer PAT, Kaiser WA. Assessing the degree of collinearity among the lesion features of the MRI BI-RADS lexicon. *Eur J Radiol* 2011;80(3):e322–e324.
- Mulaik SA. *Foundations of Factor Analysis*. 2nd ed. Boca Raton, Fla: Chapman & Hall/CRC, 2009.
- van der Velden BHM, Elias SG, Bismeijs T, et al. Complementary Value of Contralateral Parenchymal Enhancement on DCE-MRI to Prognostic Models and Molecular Assays in High-risk ER-/HER2- Breast Cancer. *Clin Cancer Res* 2017;23(21):6505–6515.
- Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15(2):R29.
- Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 2007;25(10):1239–1246.
- Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med* 2009;360(8):790–800.
- Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989;63(1):181–187.
- Pelletier J, Thomas G, Volarević S. Ribosome biogenesis in cancer: new players and therapeutic avenues. *Nat Rev Cancer* 2018;18(1):51–63 [Published correction appears in *Nat Rev Cancer* 2018;18(2):134].
- Lempiäinen H, Shore D. Growth control and ribosome biogenesis. *Curr Opin Cell Biol* 2009;21(6):855–863.
- Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490(7418):61–70.
- Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer* 2016;16(9):582–598.
- Li H, Zhu Y, Burnside ES, et al. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of MammaPrint, Oncotype DX, and PAM50 gene assays. *Radiology* 2016;281(2):382–391.
- Reyal F, van Vliet MH, Armstrong NJ, et al. A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. *Breast Cancer Res* 2008;10(6):R93.
- Costouros NG, Lorange D, Zhang Y, et al. Microarray gene expression analysis of murine tumor heterogeneity defined by dynamic contrast-enhanced MRI. *Mol Imaging* 2002;1(3):301–308.
- 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.
- Mazurowski MA, Zhang J, Grimm LJ, Yoon SC, Silber JL. Radiogenomic analysis of breast cancer: luminal B molecular subtype is associated with enhancement dynamics at MR imaging. *Radiology* 2014;273(2):365–372.
- Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5(1):4006.
- Jansen RW, de Jong MC, Kooi IE, et al. MR Imaging Features of Retinoblastoma: Association with Gene Expression Profiles. *Radiology* 2018;288(2):506–515.
- Casasent AK, Schalck A, Gao R, et al. Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing. *Cell* 2018;172(1–2):205–217.e12.