

## Article

# Identifying Associations between DCE-MRI Radiomic Features and Expression Heterogeneity of Hallmark Pathways in Breast Cancer: A Multi-Center Radiogenomic Study

Wenlong Ming <sup>1,2,†</sup>, Yanhui Zhu <sup>3,†</sup>, Fuyu Li <sup>1</sup>, Yunfei Bai <sup>1</sup> , Wanjun Gu <sup>1,4</sup>, Yun Liu <sup>5</sup>, Xiao Sun <sup>1</sup> , Xiaolan Liu <sup>3,\*</sup> and Hongde Liu <sup>1,\*</sup> 

- <sup>1</sup> State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China
  - <sup>2</sup> Division of Medical Image Computing, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
  - <sup>3</sup> Department of Breast Surgery, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China
  - <sup>4</sup> Collaborative Innovation Center of Jiangsu Province of Cancer Prevention and Treatment of Chinese Medicine, School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing 210023, China
  - <sup>5</sup> Department of Information, The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing 210029, China
- \* Correspondence: liuxiaoan@njmu.edu.cn (X.L.); liuhongde@seu.edu.cn (H.L.)  
† These authors contributed equally to this work.



**Citation:** Ming, W.; Zhu, Y.; Li, F.; Bai, Y.; Gu, W.; Liu, Y.; Sun, X.; Liu, X.; Liu, H. Identifying Associations between DCE-MRI Radiomic Features and Expression Heterogeneity of Hallmark Pathways in Breast Cancer: A Multi-Center Radiogenomic Study. *Genes* **2023**, *14*, 28. <https://doi.org/10.3390/genes14010028>

Academic Editors: Rujiao Li, Yan Zhang, Jianzhong Su and Albert Jeltsch

Received: 3 November 2022

Revised: 12 December 2022

Accepted: 20 December 2022

Published: 22 December 2022



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Background: To investigate the relationship between dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) radiomic features and the expression activity of hallmark pathways and to develop prediction models of pathway-level heterogeneity for breast cancer (BC) patients. Methods: Two radiogenomic cohorts were analyzed ( $n = 246$ ). Tumor regions were segmented semiautomatically, and 174 imaging features were extracted. Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were performed to identify significant imaging-pathway associations. Random forest regression was used to predict pathway enrichment scores. Five-fold cross-validation and grid search were used to determine the optimal preprocessing operation and hyperparameters. Results: We identified 43 pathways, and 101 radiomic features were significantly related in the discovery cohort ( $p$ -value  $< 0.05$ ). The imaging features of the tumor shape and mid-to-late post-contrast stages showed more transcriptional connections. Ten pathways relevant to functions such as cell cycle showed a high correlation with imaging in both cohorts. The prediction model for the mTORC1 signaling pathway achieved the best performance with the mean absolute errors (MAEs) of 27.29 and 28.61% in internal and external test sets, respectively. Conclusions: The DCE-MRI features were associated with hallmark activities and may improve individualized medicine for BC by noninvasively predicting pathway-level heterogeneity.

**Keywords:** radiogenomics; association analysis; breast cancer; DCE-MRI; pathway; radiomics; machine learning

## 1. Introduction

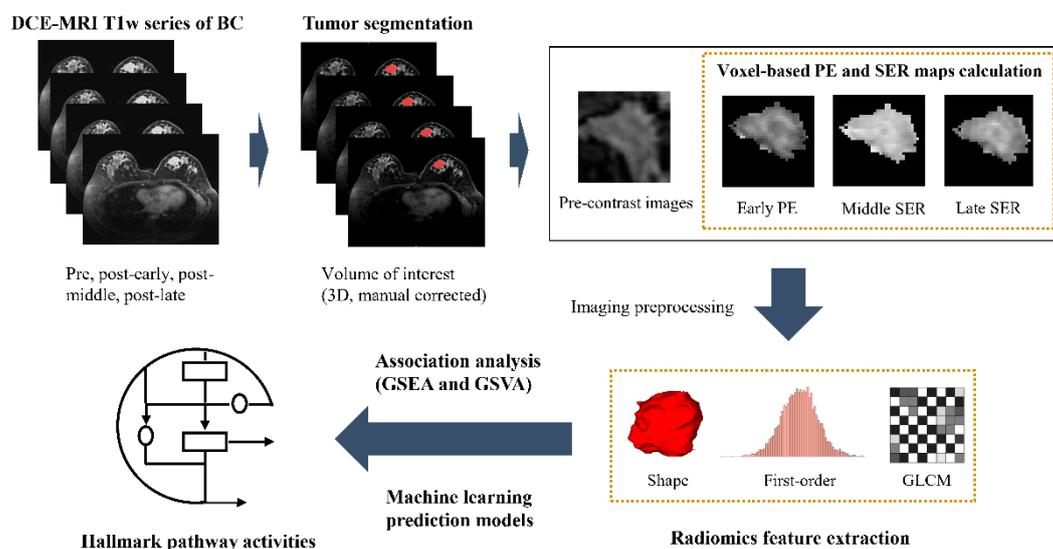
Breast cancer (BC) is currently the most frequent malignant tumor worldwide (11.7% of new tumors in 2020), with an estimated 2.3 million new cases, and the death rates of BC were 15.0 and 12.8 per 100,000 cases in developing and developed countries, respectively, indicating that BC is a serious health threat to women [1]. The high degree of heterogeneity poses a challenge to accurate diagnosis and the decision of treatment strategy in the clinical practices of BC [2,3]. Based on the expression profiling of 50 important genes,

BC can be categorized into five major molecular subtypes, including luminal-A, luminal-B, HER2-enriched, basal-like, and normal-like, which is also known as the prediction analysis on microarray (PAM50) intrinsic molecular subtypes [4,5]. Different PAM50 subtypes of BC have distinct risks of recurrence, metastasis, and response to treatment [6,7]. Accurate and timely identification of the intrinsic molecular subtypes of patients is essential for the clinical management of BC. However, due to the high cost of RNA-sequencing, immunohistochemistry (IHC) testing is the commonly used alternative method in clinics to identify the molecular characteristics of BC by detecting the gene expression of four specific biomarkers, including the estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki67 [8,9]. These findings on the transcriptomic heterogeneity of BC, including both the intrinsic expression pattern and status of the IHC biomarker, have greatly contributed to precision medicine for BC and improved overall patient outcomes [10,11]. Most of these studies were based on a single gene or a small number of genes and have been highly successful. The expression heterogeneity at the level of the pathway is also considered a very important feature of BC, but a lot of work is still needed in this area [12–14]. For example, the hallmark pathways are derived from various overlapping original gene sets, which can reduce noise and redundancy and provide a better delineated biological space to characterize the transcriptomics of disease [15]. Non-invasive access to the heterogeneity at the hallmark pathway level might contribute to personalized medicine for BC.

In the last decade, with the advancement of data science and artificial intelligence technologies, radiomics has been rapidly developed within the field of oncology research, including BC [16,17]. Identifying the reliable associations between quantitative medical imaging features and vital information about the disease, such as genetic mutations, molecular or clinical features, and treatment responses, as well as non-invasively predicting these cancer characteristics from imaging features are the core goals of radiomic research. As a subset of radiomics, radiogenomics focuses on analyzing the association of tumor genomic or transcriptomic characteristics with imaging features [18,19]. The association between transcriptomic characteristics and imaging features has been investigated in BC [20–24], for example, the distinct expression activity of cell cycle pathways was related to the imaging characteristics and clinical outcomes of BC patients [21]. However, the current studies have some shortcomings. First, most of the associations between the transcriptomic characteristics and imaging features in previous BC radiogenomics studies have been reported in a single dataset only, lacking the validation of an independent dataset. Second, few studies have analyzed the association between pathway heterogeneity and imaging features, especially for the representative hallmark gene sets. Finally, to our knowledge, no radiomic feature-based machine learning prediction models of pathway expression activity in individual BC patients have been built and reported.

In this work, to explore the associations between the imaging features of BC patients and the transcriptional heterogeneity of the hallmark pathways, two radiogenomics cohorts were collected from local and public databases. The workflow of this radiogenomic study is illustrated and detailed in Figure 1. The dynamic contrast-enhanced MRI (DCE-MRI) data and RNA-Seq data from the radiogenomics cohorts were analyzed. DCE-MRI is the most accurate modality for BC imaging now with the advantages of high resolution and temporal contiguity. Using contrast agents, DCE-MRI can amplify the differences between tumor tissue and normal tissue, which helps to reflect the imaging characteristics of tumors. The discovery cohort, including 174 patients, was used to identify the associations, and the multi-center validation cohort, including 72 patients, was used to validate. Then, the preference for significant transcriptomic–radiomic correlations was further analyzed and reported for the first time. Namely, different types of imaging features were associated with diverse pathway activities, and, correspondingly, hallmark pathways with various biological functions displayed distinct imaging correlations. Finally, we developed machine learning regression models for the enrichment scores (ESs) of some key hallmark pathways in BC based on the DCE-MRI radiomic features in an attempt to non-invasively predict

pathway expression heterogeneity in individual patients. We hope this exploratory work can contribute to precision medicine in BC.



**Figure 1.** The workflow of this DCE-MRI- and RNA-Seq-data-based radiogenomics study.

## 2. Materials and Methods

### 2.1. Subjects and Pathology Evaluation

This study was approved by the local institutional ethics committee. Written informed consent was obtained for each local patient because patients had been informed that their data, including specimens and images, might be used for possible research in the future. We employed two independent radiogenomic cohorts from multiple medical centers. The discovery cohort was collected from the local database between August 2016 and December 2018 and included 174 BC patients whose preoperative T1-weighted DCE-MRI and tumor specimens were both available. Data of DCE-MRI and RNA-seq of the external validation cohort were retrieved from the Cancer Imaging Archive (TCIA) and the Cancer Genome Atlas (TCGA) databases from four centers, including the Memorial Sloan Kettering Cancer Center, Mayo Clinic, University of Pittsburgh Medical Center, and Roswell Park Cancer Institute [25]. The detailed inclusion and exclusion criteria of both cohorts are displayed in Figure S1, and finally, we obtained and analyzed 174 and 72 patients in the discovery and validation cohorts, respectively.

ER, PR, HER2, and Ki67 were used to determine the clinical immunohistochemistry (IHC) subtypes for each patient in the discovery cohort. ER-positive, HER2-negative, high-PR expression (more than 20%), and low-Ki67 expression (less than 20%) subjects were defined as luminal-A. ER-positive, HER2-negative, low-PR, or high-Ki-67 expression subjects were defined as luminal-B. In addition, ER- and HER2-positive subjects were defined as luminal-B. ER, PR-negative, and HER2-positive patients were HER2-positive, and finally, all negative samples were triple-negative BC (TNBC).

### 2.2. DCE-MR Imaging

The DCE-MR images of the discovery cohort were scanned and acquired using Siemens TrioTim 3-Tesla (3T) scanner (Siemens Healthcare, Erlangen, Germany) in the axial position. Gadolinium-diethylenetriamine pentaacetic acid (Gd-DTPA) was used as the contrast agent, and patients were injected with the agent at an amount of 15 mL in a dose of 0.1 mmol/kg. The dynamic T1-weighted MRI sequences with fat suppression were acquired at six-timepoints, including one pre-contrast and five post-contrast (time interval was approximately 1 min). For most images from the discovery cohort, the echo time was 15.7 ms, the repetition time was 423 ms, the flip angle was 10 degrees, the field of view was 340 × 340 mm, the matrix size was 448 × 448 pixels, and the slice thickness was 0.9 mm.

For the multi-center validation cohort, using the T1-weighted 3D spoiled gradient-echo sequence, GE scanners (GE Healthcare, Chicago, IL, USA) on a 1.5 Tesla magnet strength scanned and generated the images. The in-plane resolution of validation images ranged from 0.53 to 0.86 mm, spacing between slices ranged from 2 to 3 mm, the flip angle was 10 degrees, and the acquisition matrix was  $256 \times 192$ .

### 2.3. Segmentation of Tumor Regions

Before extracting the quantitative radiomic features, tumor segmentation, and image pre-processing were conducted. Firstly, based on the differences in kinetic characteristics between tumor tissue and normal tissue during dynamic contrast enhancement, we segmented and obtained the rough volume of interest (VOI) of the tumor mask by setting a threshold of intensity value on the registered and subtracted images of the first enhanced MRI sequences in the open-source software 3D Slicer (version 4.8.1, Cambridge, MA, USA). Secondly, the tumor masks were corrected manually and finally confirmed by two radiologists blinded to the clinical information, one with 10 years and another with 3 years of breast imaging experience.

### 2.4. Calculation of Percentage Enhancement and Signal Enhancement Ratio Maps

By taking advantage of the DCE-MRI technique, we defined and calculated the voxel-based percentage enhancement (PE) and signal enhancement ratio (SER) maps based on the signal intensity of each voxel for each patient in this work [23,26,27]. Firstly, the N4 bias correction algorithm was used to reduce the possible heterogeneity bias of 3T MR images [28]. Then, with the purpose of extracting the dynamic radiomic features of signal intensity in the tumor region after contrast agent injection, MRI sequences at four timepoints, including pre-enhanced (scan before contrast injection), early, middle, and late post-enhanced images (scans at approximately 1, 3, and 5 min after contrast injection), were selected and used to calculate the intensity ratio maps. The early PE map for each patient was acquired using Equation (1):

$$PE_{early} = 100 \times \frac{I_{early} - I_{pre}}{I_{pre}} \quad (1)$$

where  $I_{early}$  was the signal intensity of each voxel in early enhancement, and  $I_{pre}$  was the voxel initial intensity before contrast. The SER maps of both middle and late enhancement were calculated using Equation (2):

$$SER_{map} = \frac{I_{early} - I_{pre}}{I_{map} - I_{pre}} \quad (2)$$

where  $I_{map}$  here is the voxel-enhanced signal intensity in middle or late images. After that, we performed the subsequent radiomic feature extraction using pre-contrast MR images and early PE, middle, and late SER maps. The map calculation was conducted in Python 3.5.2 (Virginia, USA).

### 2.5. DCE-MRI Radiomic Feature Extraction

An open-source Python package, pyradiomics (version 2.2.0), was used to perform image normalization and feature calculation [29]. Image normalization included remapping the histogram to fit within  $\mu \pm 3\sigma$  ( $\mu$  means gray level within the volume of segmentation;  $\sigma$  means gray-level standard deviation) and resampling images to an isotropic voxel resolution of 1 mm using the B-spline method. Finally, according to the image biomarker standardization initiative (IBSI), 174 quantitative radiomic features, including 14 tumor-shape features, 72 histogram features, and 88 gray-level co-occurrence matrix (GLCM) features, were calculated for each patient. All procedures of image pre-processing and feature extraction were completed in Python 3.5.2 (Virginia, USA).

## 2.6. RNA-Sequencing and Bioinformatics Analysis

Tumor tissue of 199 BC patients was collected in the local database. The total RNA was obtained using VAHTS Total RNA-seq (H/M/R) Library Prep Kit for Illumina with the manufacturer protocol and stored in liquid nitrogen at  $-80^{\circ}\text{C}$ . Ovation human FFPE RNA-seq library systems were used to construct RNA-seq libraries, and RNA was sequenced on an Illumina HiSeq X Ten platform using paired-end 150 bp runs. We used Trimmomatic to control the sequencing quality [30]. Then, the RNA-seq reads were aligned to human genome 19 using STAR29 and quantified using HTSeq-Count [31,32]. Finally, expression values of 57,773 transcripts were quantified in the forms of both counts and FPKM (fragments per kilobase of exon per million reads mapped) for each patient in the discovery cohort.

Both the qualitative and quantitative associations between 174 imaging features and the 50 hallmark pathways (defined in <http://www.gsea-msigdb.org/gsea/msigdb/human/genesets.jsp?collection=H>, accessed on 20 January 2020) were investigated in this work. To obtain the qualitative associations, patients were classified into high and low groups for each imaging feature using the median as the cutoff. The gene differential expression analysis between the high and low groups was then conducted using DESeq2 [33]. After that, the gene set enrichment analysis (GSEA) was performed to identify the hallmark gene sets that were significantly enriched in the high- or low-imaging feature value groups by determining the positions of the genes of a hallmark pathway in the ranked all gene list based on differential expression analysis. The P values of hallmark–imaging correlation pairs were corrected using false discovery rate (FDR), and an FDR-corrected P value less than 0.05 could finally be retained [34]. The quantitative associations were calculated using the gene set variation analysis (GSVA) and Pearson correlation coefficient. GSVA was performed using the R package GSVA based on the RNA-seq data to obtain the enrichment scores (ESs) of pathways in individuals [35], and Pearson correlation coefficient was used to calculate the relationships between radiomic features and ESs.

## 2.7. Feature Pre-Processing and Machine Learning Modeling

To non-invasively predict the enrichment scores (ESs) of GSVA of hallmark pathways from radiomic features, the random forest (RF) regression-based models were trained and tested using the scikit-learn Python package (version 1.0.2). The radiogenomics discovery cohort was randomly divided into an internal training set ( $n = 139$ ) and an internal independent test set ( $n = 35$ ) according to 80 and 20%, and the validation cohort was used as the external test set ( $n = 72$ ). In this work, optimization was performed for both feature pre-processing and model hyper-parameters. Five pre-processing operations were considered before training the model. The first operation was whether to normalize the image features using the Z-score method. The second one was whether to perform oversampling to remove the imbalance of the sample distribution. The specific process was performed by discrete the training set samples into 10 bins according to the enrichment scores of the target pathway and then balancing the number of samples in each interval by oversampling. The third operation was whether to construct the final training model using only the 15 most important image features determined by the 5-fold cross-validation training process. The fourth operation was whether to set weights on the samples based on the enrichment scores of the target pathway to balance the sample inhomogeneity. The final operation was whether to optimize the RF hyper-parameters automatically or just use the default hyper-parameters. If the model should be optimized, five important hyper-parameters of RF, including the number of trees, node criterion, max depth, max samples, and max features were optimized using grid search with 5-fold cross-validation and negative mean absolute error. The details of the RF parameters to be optimized in this work were as follows: `n_estimators`: [30, 50, 100, 250, 500, 1000, 2000]; `criterion`: [“absolute\_error”, “squared\_error”, “poisson”]; `max_depth`: [2, 4, 6, 8, 10, none]; `max_samples`: [0.2, 0.4, 0.6, 0.8, none]; and `max_features`: [0.2, 0.4, 0.6, 0.8, 1.0]. Five-fold cross-validation

(CV) was applied to avoid overfitting. The pre-processing and machine learning analysis was performed in Python 3.8.10 (Virginia, USA).

### 2.8. Statistical Analysis

Two-sided Pearson's chi-squared test or Fisher's exact test were used to determine whether there was a significant difference between the two cohorts. The quantitative association between radiomic features and the GSVA enrichment scores of the pathways was given by the Pearson correlation coefficient. *P*-values smaller than 0.05 were considered significant. Mean absolute error (MAE) was used to evaluate the prediction performance of RF prediction models, and the 95% confidence interval (CI) was used to evaluate the performance stability during the 5-fold CV. The statistical analysis was performed using R 4.0.1 (Murray Hill, NJ, USA).

## 3. Results

### 3.1. Broadly Significant Associations between Hallmark Pathways and DCE-MRI Radiomic Features in BC

We performed statistical analysis and the presentation of both the demographic information of the radiogenomic cohorts and the significantly associated imaging features or hallmark pathways based on GSEA, respectively. Table 1 details the clinical characteristics of the two radiogenomic cohorts. Except for the PAM50 molecular subtypes, the two datasets had no differences in other characteristics. For the GSEA-based association analysis, we observed 101 radiomic features, and 43 hallmark pathways were significantly associated with GSEA (*p*-value < 0.05) in the discovery cohort, and 138 radiomic features and 41 hallmark pathways in the validation cohort (Table 2). The number of all of the significant qualitative association pairs of the two cohorts were 805 and 1114 (accounting for 9.3 and 12.8% of all possible pathway-feature combinations), respectively. Although the absolute proportions of significant associations were low, more than 80% of hallmark pathways in both cohorts (86 and 82%, respectively) showed significant imaging feature correlations (Table 2), demonstrating the widespread impact of the transcriptional activity on the phenotypes of complex disease. In addition, the proportion of radiomic features significantly associated with the hallmark pathways to all of the imaging features was 58 and 79.3%, respectively (Table 2). For a hallmark pathway, an average of 18.7 and 27.2 radiomic features were associated with it, respectively. Conversely, for a radiomic feature, there were averagely 7.9 and 8.1 hallmark pathways associated with it.

**Table 1.** The clinical and molecular characteristics of BC tumors in the radiogenomic discovery and validation cohorts.

Characteristics	Discovery Cohort (n = 174)	Validation Cohort (n = 72)	p-Value
Age, mean (SD)	≤50y: 95/>50y: 79; 49.78y (9.99)	≤50y: 30/>50y: 42; 53.96y (11.75)	0.088 <sup>a</sup>
<b>IHC receptors</b>			
ER status	P:127/N:47	P:61/N:11	0.071 <sup>a</sup>
PR status	P:111/N:63	P:55 N:17	0.077 <sup>a</sup>
HER2 status	P:36/N:138	P:14 N:37/NA:21	0.407 <sup>a</sup>
Ki67 status	high:136/low:38	NA	NA
<b>IHC-based subtype</b>			
Luminal-A	28	NA	NA
Luminal-B	101	NA	NA
HER2-positive	15	NA	NA
Triple-negative	30	NA	NA
<b>PAM50 subtype</b>			
Luminal-A	49	44	<0.001 <sup>b</sup>
Luminal-B	43	9	
HER2-enriched	29	5	
Basal-like	43	10	
Normal-like	10	4	
<b>Pathological stage</b>			
Stage I	55	17	0.267 <sup>a</sup>
Stage II	94	47	
Stage III	25	8	

Note: unless otherwise indicated, data are the number of patients or P-value of the statistical test. <sup>a</sup> p-value for the two-sided Pearson's chi-squared test, <sup>b</sup> p-value for the two-sided Fisher's exact test. P for positive and N for negative. NA: not available.

**Table 2.** Percentages of hallmark pathways or imaging features with radiomic–transcriptomic associations. The proportion of pathways with significant imaging correlation with all 50 hallmark pathways, and the proportions of DCE-MRI features with significant transcriptomic association with all imaging features of that type were calculated and displayed for both cohorts.

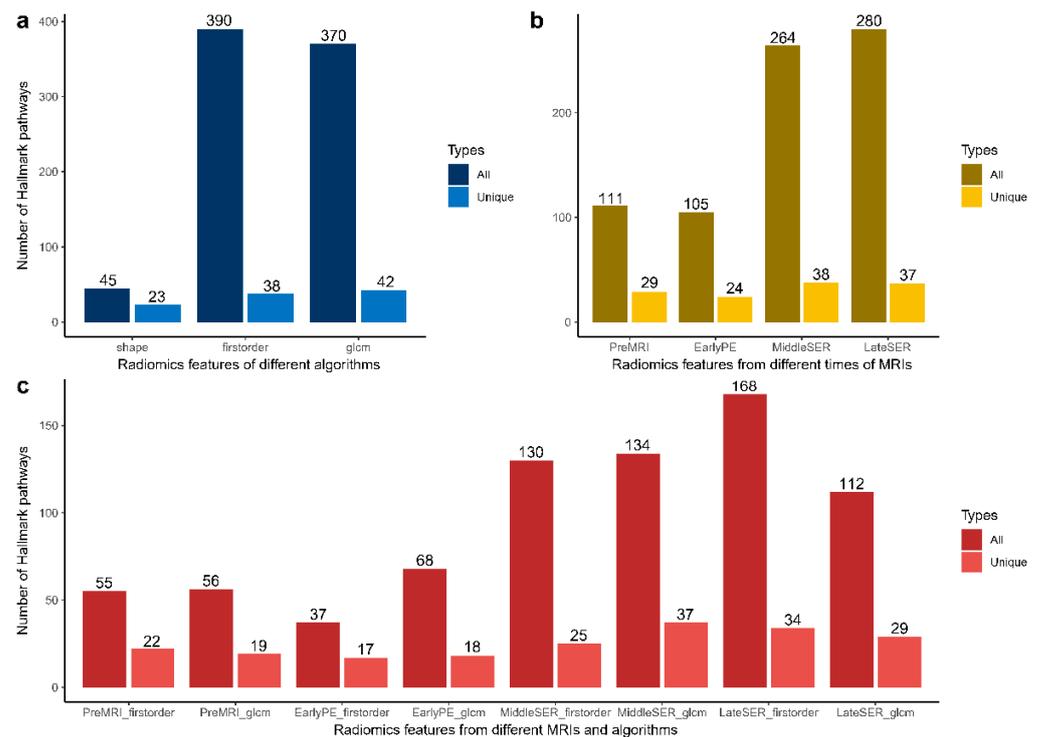
Pathways or Features	Discovery Cohort	Validation Cohort
Unique associated hallmark pathways	43 (86.0%)	41 (82.0%)
Unique associated DCE-MRI radiomic features	101 (58.0%)	138 (79.3%)
Shape features	9 (64.3%)	14 (100.0%)
First-order features	46 (63.9%)	56 (77.8%)
GLCM features	46 (52.3%)	68 (77.3%)
Pre-MR images	18 (11.3%)	33 (18.3%)
Early PE images	21 (13.1%)	34 (21.3%)
Middle SER images	25 (15.6%)	26 (17.5%)
Late SER images	28 (16.3%)	31 (19.4%)

Note: unless otherwise indicated, data are the number of hallmark pathways or DCE-MRI features which showed a statistically significant radiomic–transcriptomic association. The percentages of significant pathways or features to the total number of pathways or features of the respective type to which they belong are in parentheses.

### 3.2. Radiomic Features of Tumor Shape and Mid-To-Late Enhanced Stages Showed More Transcriptional Associations

We further investigated whether imaging features related to hallmark pathway activity, identified by GSEA, differed significantly across the feature categories and reported these findings for the first time. Using the algorithm-based imaging feature category criteria (feature category I), 174 radiomic features were divided into shape, first-order, and GLCM features. The first-order and GLCM features contributed the most significant correlations, but the shape features had the highest ratio of transcriptional associations (Figure 2a,

Figure S2a, and Table 2). We then compared the radiomic features extracted from MR images at different times (feature category II) and found that features from the SER images, at the time of late contrast enhancement, showed relatively high transcriptional associations in both cohorts (Figure 2b, Figure S2b, and Table 2). This temporal preference persisted when considering the features of different algorithms and images (feature category III), except for the validation cohort where PE images from early enhancement had more transcriptional associations (Figures 2c and S2c). Furthermore, due to the possible one-to-multiple relationships, the number of unique radiomic-associated hallmark pathways of each type of DCE-MRI feature was calculated (Figure 2, Figure S2, and Table 3). The two-sided chi-squared test showed that there was no significant difference in the transcriptional association between different categories of imaging features in the two cohorts.



**Figure 2.** GSEA-based associations of different kinds of DCE-MRI features with hallmark pathways in the discovery cohort of BC radiogenomics. Imaging features were categorized into different classes according to three different definitions. The number of significant associations between different algorithm imaging features and hallmark pathways are displayed in (a), the number of significant associations between imaging features at different times during enhancement and hallmark pathways are shown in (b), and (c) presents the numbers of imaging features significantly related to hallmark pathways for different algorithms and at different times. As an imaging feature may be significantly associated with more than one hallmark pathway based on the GSEA, not only the absolute number of all significant associations was counted (type: All) but also the number of unique hallmark pathways was obtained by removing the duplicated pathways (type: Unique). Numbers represent the number of significant associations.

**Table 3.** The number of radiomic-associated hallmark pathways of each type of DCE-MRI feature. The number of significant imaging-correlated hallmark pathways was counted for each class of imaging features in both cohorts, and whether the radiomic–transcriptomic associations were significantly different in the two independent datasets is shown.

Characteristics	Discovery Cohort	Validation Cohort	<i>p</i> -Value
<b>Feature category I</b>			0.801
Shape features	23	27	
First-order features	38	37	
GLCM features	42	39	
<b>Feature category II</b>			0.324
Pre-MR images	29	29	
Early PE images	24	36	
Middle SER images	38	31	
Late SER images	37	32	
<b>Feature category III</b>			0.129
First-order from pre-MR images	22	25	
GLCM from pre-MR images	19	26	
First-order from early PE images	17	27	
GLCM from early PE images	18	34	
First-order from middle SER images	25	30	
GLCM from middle SER images	37	26	
First-order from late SER images	34	30	
GLCM from late SER images	29	22	

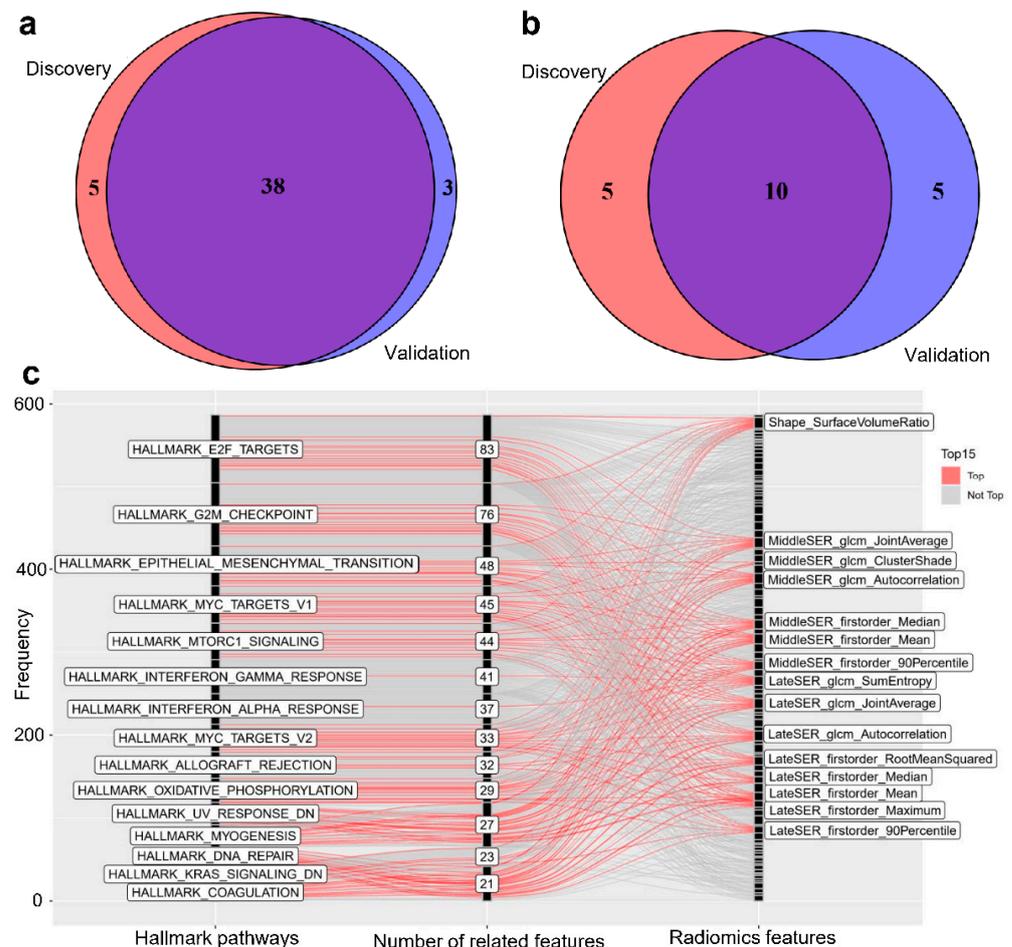
Note: unless otherwise indicated, data are the number of unique hallmark pathways or the *p*-value of the two-sided Pearson’s chi-squared test.

### 3.3. The Significant Radiomic-Associated Hallmark Pathways Demonstrated Molecular Function Preferences

On the other hand, we attempted to identify qualitative GSEA-based imaging–transcriptomic associations that were stable across different datasets and uncover whether there is an overlap in the biological function between the hallmark pathways that have significant correlation with imaging features. The results from the two independent radiogenomic cohorts demonstrated a significant and stable relationship between the hallmark pathways and radiomic features (Figures 3 and S3). The intersection of the number of pathways with the radiomic–transcriptomic association in the two cohorts was 38 (Figure 3a), and the intersection of the DCE-MRI radiomic features was 81 (Figure S3b). We first ranked the hallmark pathways by the number of significant associations with the radiomic features and selected the top 15 pathways in both radiogenomic cohorts. Ten hallmark pathways, including “E2F TARGETS”, “G2M CHECKPOINT”, “MYC TARGETS V1”, “MTORC1 SIGNALING”, “KRAS SIGNALING DN”, “EPITHELIAL MESENCHYMAL TRANSITION”, “MYOGENESIS”, “INTERFERON GAMMA RESPONSE”, “INTERFERON ALPHA RESPONSE”, and “ALLOGRAFT REJECTION”, showed broad and stable imaging phenotype associations in both radiogenomic cohorts (Figures 3b–c and S3a). For example, the pathway “E2F TARGETS” was remarkably related to 83, and 84 DCE-MRI features in the discovery and validation cohorts, ranking first and third, respectively (Figures 3c and S3a). Although imaging features significantly associated with pathway activity also overlap remarkably in both datasets, there was a notable difference in the ranking of the transcriptional associations of the DCE-MRI radiomic features in the two cohorts (Figure S2b–c). The first-order and GLCM features extracted from the middle and late SER images exhibited a high proportion of transcriptional correlations in the discovery cohort, yet in the validation cohort, the features of early PR images and tumor shape ranked high (Figures 3c and S3a). This may be explained by the high system bias of the MRI data in the validation cohort.

In order to further explore the heterogeneity of imaging-related transcriptional activities, the 10 remarkable pathways were grouped into 4 categories based on their molecular functions. Cell cycle-related pathways included “E2F TARGETS” and “G2M CHECK-

POINT". Proliferation-related pathways consisted of "MYC TARGETS V1", "MTORC1 SIGNALING", and "KRAS SIGNALING DN". Extracellular matrix (ECM)-related pathways included "EPITHELIAL MESENCHYMAL TRANSITION" and "MYOGENESIS". Additionally, "INTERFERON GAMMA RESPONSE", "INTERFERON ALPHA RESPONSE", and "ALLOGRAFT REJECTION" attributed to the immune-related pathways. The four major categories were further analyzed using GSVA and demonstrated in our subsequent results.

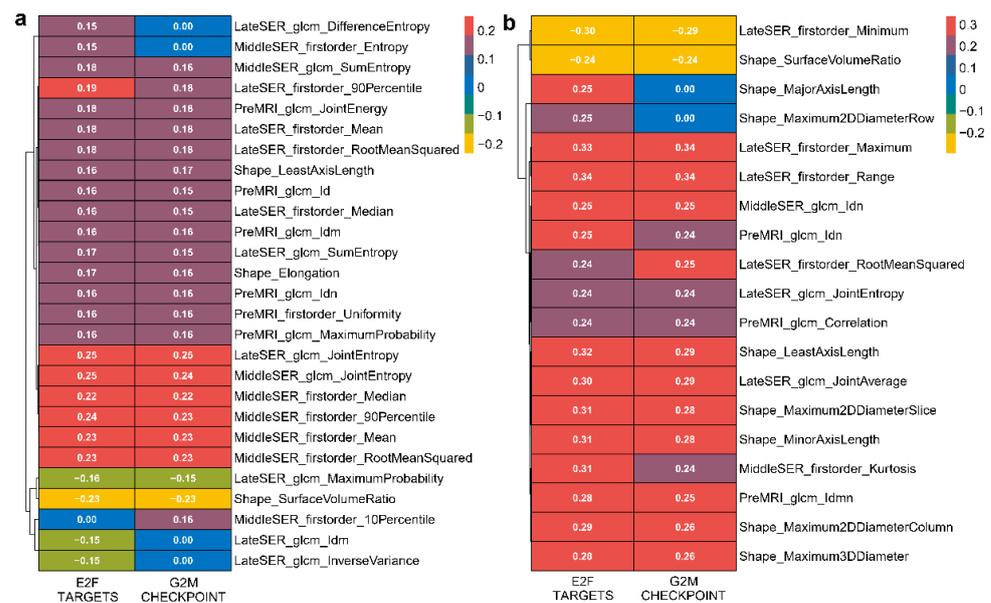


**Figure 3.** Hallmark pathways with significant imaging correlation were highly overlapping in both cohorts. The distribution of the pathways related to imaging features in the discovery and validation cohorts, shown in (a,b), presents the overlap of the top 15 pathways in the discovery dataset and the top 15 pathways in the validation dataset, as ordered by the number of significant associations. The details of the top 15 hallmark pathways and the top 15 DCE-MRI features of the discovery cohort are presented in (c), and the numbers on the middle axis represent the number of significant associations.

### 3.4. The Association Heterogeneity between the Four Categories of Hallmark Pathways and Radiomic Features

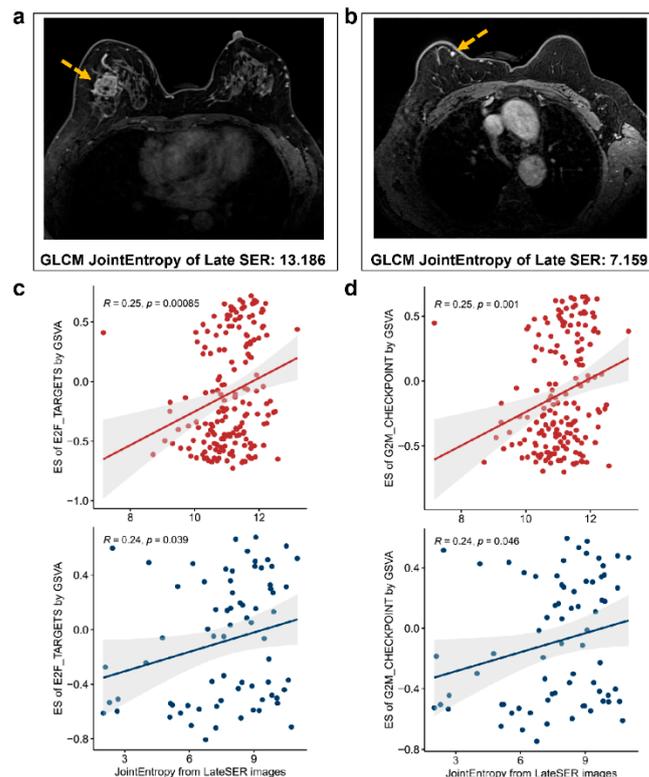
To quantify the radiomic–transcriptomic correlation, we calculated and reported the Pearson correlation coefficient of each hallmark pathway enrichment score obtained using GSVA with each DCE-MR imaging feature value. Firstly, many significant positive Pearson correlations between cell cycle-related pathways and imaging features were observed in both the discovery and validation cohorts (Figure 4). It was noteworthy that the GLCM feature JointEntropy from the late SER map had the highest Pearson correlation coefficient values of 0.25 and 0.25 in the discovery cohort and 0.24 and 0.24 in the validation cohort with the expression variation scores of cell cycle-related pathways "E2F TARGETS" and "G2M CHECKPOINT", respectively (Figure 4a,b). We further illustrated the differences

in imaging between patients with high and low late SER JointEntropy (Figure 5a,b) and revealed that patients with higher late SER JointEntropy showed significantly higher enrichment scores in the cell cycle-related pathways, such as “E2F TARGETS” and “G2M CHECKPOINT” ( $p$ -value < 0.05, Figure 5c,d). Our results demonstrate that the relationship between the late SER JointEntropy and the two hallmark pathways exhibited a high degree of consistency and robustness in both datasets. In addition, the tumor-shape feature SurfaceVolumeRatio exhibited the greatest negative relationships with the GSVA scores of “E2F TARGETS” and “G2M CHECKPOINT”, as the Pearson correlation coefficients ( $p$ -value < 0.05) of  $-0.23$  and  $-0.23$  were found for the discovery cohort and  $-0.24$  and  $-0.24$  for the validation cohort, respectively (Figure 4a,b).



**Figure 4.** Quantitative associations of GSVA enrichment scores of two cell cycle-related hallmark pathways with imaging features. The significant Pearson correlation coefficients of the enrichment scores of both “E2F TARGETS” and “G2M CHECKPOINT” pathways with the imaging features for the discovery and validation cohorts are detailed in (a,b), respectively. Non-zero values in the cells represent significant correlation coefficients, and zero represents non-significant correlation coefficients.

Three proliferation-related pathways were found related to more tumor-size features than other hallmark pathways (Figures S4a and S5a). Two tumor-size features, including LeastAxisLength and MinorAxisLength, both indicating the minimum size of the tumor, were significantly positively related to the “MTORC1 SIGNALING” pathway ( $p$ -value < 0.05). Interestingly, many significantly negative correlations between the two ECM-related pathways (“EPITHELIAL MESENCHYMAL TRANSITION” and “MYOGENESIS”) and radiomic features were revealed (Figures S4b and S5b). The GLCM feature Idn, from the pre-contrast MRI, showed the most significant negative correlation with both pathways, with the  $r$  values of  $-0.27$  and  $-0.3$ , respectively ( $p$ -value < 0.05, Figure S4b). In addition, most of these negatively correlated imaging features, such as the first-order feature Mean, Median, 90Percentile, and 10Percentile from the late SER map, were associated with tumor signal enhancement (Figures S4b and S5b). For the immune-related pathways in the discovery cohort, most of the pathway-related features were extracted from both the middle and late SER maps (Figure S4c); however, the significantly correlated features in the validation dataset were mainly from the early PE images (Figure S5c). Despite some discordant results, overall, the significant correlations between the DCE-MRI features and GSVA scores of the hallmark pathways were identified and validated for the first time.



**Figure 5.** Clinical imaging differences were associated with the expression heterogeneity of hallmark pathways. The MR image of a patient with a high joint entropy value of GLCM for late SER image is illustrated in (a), and in (b) the MR image of a patient with a low value of this feature is displayed (The yellow arrows point to the tumors). (c,d) show the scatter plots of the values of joint entropy in breast cancer patients in the discovery set and validation set versus the enrichment scores of the “E2F TARGETS” and “G2M CHECKPOINT” pathways, respectively.

### 3.5. Non-Invasive Prediction of the Enrichment Scores of GSVA of Hallmark Pathways

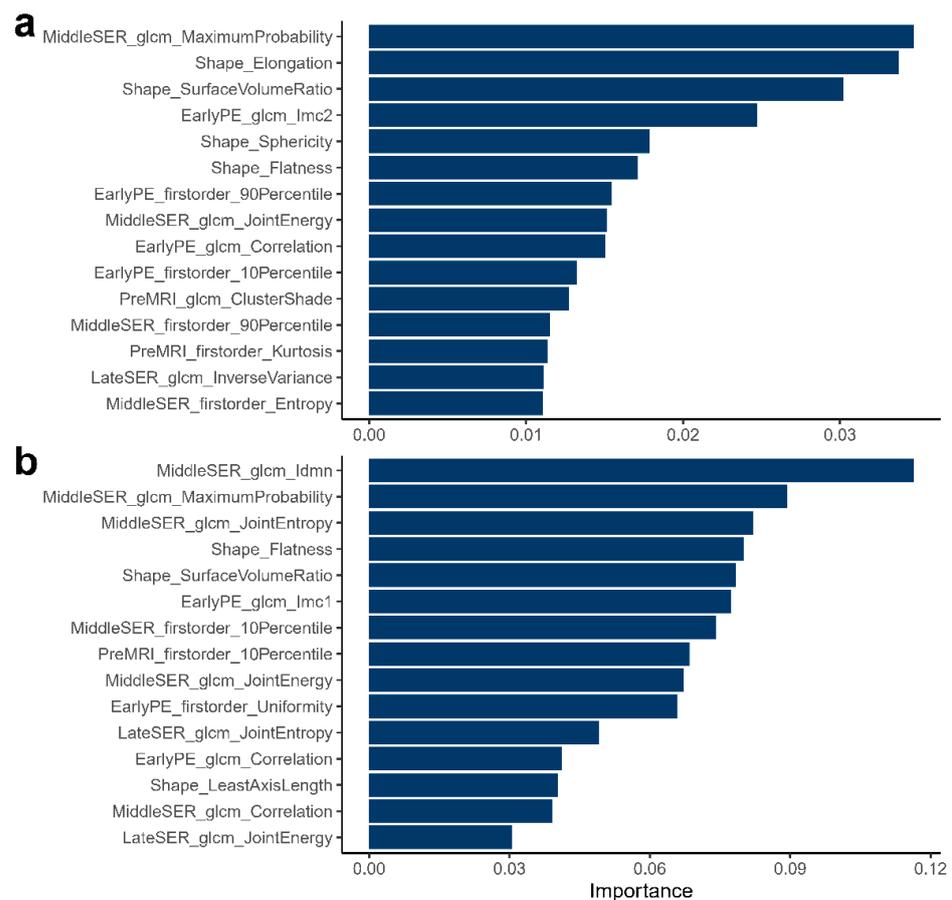
To extend the application scenario of radiomics, for the first time, in this work, we attempted to use imaging features to build machine-learning models for predicting the transcriptomic characteristics at the level of pathways in BC, namely, predicting the enrichment scores (ESs) of the GSVA in the pathways. The RF prediction models were established and validated for four representative hallmark pathways. The model was trained in the internal training set ( $n = 139$ ) and tested in the internal test set ( $n = 35$ ) and external multi-center test set ( $n = 72$ ). The training and testing MAEs of the RF regressors are reported in Table 4. The best-performing constructs of the RF models are detailed in Table S1. Compared to the other models, the RF model for the mTORC1 signaling pathway had the lowest MAEs, with 27.42% in the five-fold CV training set, and 27.29 and 28.61% in the internal and external test sets, respectively. Notably, the RF regressors for cell cycle-, proliferation-, and ECM-related pathways showed similar performances in the internal and external datasets (Table 4). However, the prediction model for the immune-related pathway (INTERFERON GAMMA RESPONSE) performed significantly worse in the external test set (40.37%) than in the internal test set (36.82%). This was consistent with the results of the GSEA-based analysis in that the association of the immune-related pathways with imaging features was more unstable. The top 15 important radiomic features for the prediction models of “G2M CHECKPOINT”, “MTORC1 signaling”, “Epithelial mesenchymal transition”, and “INTERFERON GAMMA RESPONSE” are displayed in Figures 6a–b and S6a–b, respectively. The important imaging features with predictive ability demonstrated remarkable pathway correlations. For example, the GLCM, featuring MaximumProbability from the middle SER image, was the most important feature in the RF regressor of “G2M CHECKPOINT” (Figure 6a) and significantly negatively associated with the ES of GSVA (Figure 4a). Tumor-shape characteristics such as

SurfaceVolumeRatio and Flatness were important features in all of the prediction models (Figures 6a,b and S6a,b), suggesting that tumor-shape features, as simple features that can be readily available in clinical practice, may have more application.

**Table 4.** The prediction performance of the GSVA enrichment scores of hallmark pathways based on radiomic features. The random forest prediction models of enrichment scores for four representative hallmark pathways were built in the internal training set and validated in the independent internal and external test sets. Five-fold cross-validation (CV) was performed to avoid overfitting, and MAE was used to evaluate model performance. The 95% CI was used to evaluate the performance stability during the 5-fold CV.

Hallmark Pathways	Training Set (5-Fold CV)	Internal Test Set ( $n = 35$ )	External Test Set ( $n = 72$ )
<b>G2M checkpoint</b>			
MAE	32.86% (6.5%, 59.6%)	37.43%	37.22%
<b>Epithelial mesenchymal transition</b>			
MAE	33.11% (31.2%, 35.1%)	33.04%	36.25%
<b>MTORC1 signaling</b>			
MAE	27.42% (23.3%, 31.5%)	27.29%	28.61%
<b>Interferon gamma response</b>			
MAE	35.12% (8.4%, 62.2%)	36.82%	40.37%

Note: unless otherwise indicated, data are the mean absolute errors (MAEs) for each random forest regression prediction model. Values in parentheses are 95% confidence intervals (CIs) for the MAEs of the 5-fold CV.



**Figure 6.** The top 15 important radiomic features for the random forest prediction models of representative hallmark pathways. The top 15 important DCE-MRI features of the G2M checkpoint prediction model and the mTORC1 signaling prediction model are detailed in (a,b), respectively.

#### 4. Discussion

In the present work, we investigated, for the first time, the associations of DCE-MRI radiomic features with hallmark pathway expression activities in BC based on two independent datasets. Our results indicate that more than 80% of hallmark pathways significantly correlated with imaging features in both cohorts, and at least half of the imaging features showed the relevance of transcriptional activity. Tumor-shape features and imaging features extracted from the middle and late images in the contrast enhancement exhibited broader correlations with hallmark pathway activities. The hallmark pathways associated with radiomics showed a similar preference in biological functions. Cell cycle-, proliferation-, ECM-, and immune system-related hallmark pathways have broader and more remarkable imaging associations. Moreover, we made the first attempt to develop non-invasive RF regression models of the enrichment scores of hallmark pathways in BC patients based on radiomic features, which may be useful for personalized medicine for BC.

Radiomics is an emerging field in translational medicine that is gaining increasing interest in the practice of precision medicine for a variety of cancers, such as breast cancer and prostate cancer, because of its ability to provide a promising approach to non-invasive predictions of the characterization of key gene mutation, molecular activity, progression, treatment response, and prognosis of disease [36–39]. The associations of DCE-MRI imaging features with transcriptomic activities in BC have been reported in many studies. For example, the enhancement pattern in DCE-MRI suggested the deregulation of the mTOR pathway [40], and the enhancing rim fraction score can indicate the early metastasis of BC and expression of some important long non-coding RNAs [41]. Compared to the pre-contrast and early post-contrast images, we found that radiomic features from the middle and late post-contrast images were more transcriptomic-related in the discovery cohort (Figure 2), which might be because, in the mid or late stage of dynamic contrast enhancement, the contrast agent was sufficiently absorbed and metabolized in the tumor tissue and its microenvironment to accurately reflect the imaging features of the tumor. However, features from the early post-contrast images displayed more transcriptional correlations in the validation cohort (Figure S2). This might be caused by the inconsistent acquisition time points during DCE-MRI, with the images from the local database using the newest MRI protocols, whose first post-contrast acquisition time was around 1 min, whereas the validation dataset from the public database used the old imaging protocol, whose first post-contrast adoption time point was much later. In addition, the tumor-shape features were found to be most closely connected to the hallmark transcriptional activity in both cohorts (Table 2). This result was reasonable because shape features were the most basic tumor characteristics and were affected and regulated by various transcriptional activities [22]. Although the two cohorts originated from different medical centers, the association of the imaging features with the transcriptional activity did not differ significantly between the two cohorts when the imaging features were analyzed by different types. This illustrated the robustness of the associations between imaging features and hallmark expression activities (Table 3).

Previous radiogenomic studies of BC showed that the expression activities of some genes or the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with specific functions were associated with imaging features [20,22]. In particular, KEGG pathways such as cell cycle, ECM–receptor interaction, and primary immunodeficiency have been frequently reported to be correlated with DCE-MRI imaging features [21,22]. In this work, we used hallmark gene sets instead of KEGG pathways to investigate the radiomic associations because the hallmark pathways are noise-removed and more representative. We found that most of the hallmark pathways were related to the DCE-MRI imaging features (Table 2) and that the top-ranked pathways in both cohorts had high overlap (Figure 3). By looking at the detailed biological functions involved in these notable pathways, we further found that they can be classified into four major categories, including cell cycle-, proliferation-, ECM-, and immune-related pathways. In our work, the E2F target pathways

exhibited the broadest imaging correlation, particularly a significant negative correlation with the surface-to-volume ratio and significant positive correlations with the joint entropy of the middle and late SER images (Figures 3 and 4). The E2F transcription factors regulate a wide range of genes that are essential for DNA replication and cell-cycle progression and are closely involved in the progression and metastasis of BC [42,43]. Recent research also showed that the E2F targets pathway score is a potential biomarker of a neoadjuvant therapy response in BC patients [44]. As a recently revisited hallmark of cancer development, MYC transcription factors regulate almost every aspect of cellular metabolism and are dysregulated in most types of cancers [45]. MYC can also promote tumor proliferation in BC, and it may be a potential therapeutic target of triple-negative BC patients [46]. The mTORC1 signaling pathway is a widely studied pathway with important roles in cellular protein synthesis and autophagy and is closely associated with BC metastasis and response to targeted therapies [47,48]. In addition, as an essential component of the tumor microenvironment, the ECM likewise plays an important role in mediating the progression and metastasis of BC, which may influence the process of enhanced imaging [49–51]. Playing a dual role in tumor suppression and pro-tumorigenesis in human cancers, interferon gamma is engaged in complex tumor immune regulatory mechanisms and is associated with the immune microenvironment and prognosis of BC [52–54]. Our results suggest that these hallmark gene sets, which are intimately involved in the processes of BC development, progression, metastasis, and treatment response, have clear and robust relationships to the radiomic features of DCE-MRI.

The non-invasive prediction of key molecular characteristics or clinical parameters of cancer is the core task of radiomics towards clinical application. Many efforts have been made to predict the receptor status, clinical or molecular subtypes, and other important disease features in BC patients [24,55–57]. However, few studies have focused on the non-invasive prediction of heterogeneity at the pathway level, and the vast majority of these studies were binary classification or multi-classification tasks, which are still insufficient for individualized medicine where higher precision is required. In this work, we made a preliminary attempt using DCE-MRI radiomic features to predict the enrichment scores of hallmark pathways for BC individuals. Although the random forest regression models showed poor prediction performance in both the internal and external test sets overall, the model of the mTORC1 signaling pathway performed reasonably in predicting the enrichment scores, with MAEs of 27.29 and 28.61% for the internal and external test sets, respectively (Table 4). Compared to the other three pathways, the prediction model for the interferon gamma response pathway performed significantly worse in the external dataset than in the internal dataset, probably due to the high heterogeneity of the individual immune system. Moreover, the important imaging features with predictive ability selected by RF were found to be highly overlapping with the significantly associated features obtained based on GSEA and GSVA (Figure 6 and Figure S6), which may improve the interpretability of the machine-learning models. It was also to be noted that the models did not perform very well during the training process of the five-fold CV (Table 4). The results might indicate that the imaging features we used in this work were not yet representative and that the accurate prediction of the expression heterogeneity for a specific pathway was difficult.

There were some limitations to this work. Firstly, the external cohort had a small sample size and was collected from multiple medical centers. Although we performed the required data normalization preprocessing and validated some radiomics–transcriptomics correlations in the cohort, a more refined image preprocessing and normalization approach is needed to better address data bias in a larger dataset. Secondly, we simply demonstrated and validated the association of DCE-MRI radiomic features with hallmark pathway activities in BC and analyzed the heterogeneity of this association. How to more effectively integrate these findings into clinical practice needs to be further explored in the future. Finally, the performance of the pathway enrichment score prediction models based on DCE-MRI radiomic features was still low, and the models can be further optimized by

collecting larger study cohorts and using more radiomic features or deep features based on deep learning algorithms in future works.

## 5. Conclusions

In this work, the associations of DCE-MRI radiomic features with hallmark pathway expression activities were investigated and validated in two BC radiogenomic cohorts. There was a broad and significant relationship between imaging features and hallmark expression heterogeneity. The imaging features of tumor shape and mid-to-late post-contrast stages showed more transcriptional connections. Ten hallmark gene sets of cell cycle-, proliferation-, ECM-, and immune-related pathways exhibited more significant and tighter correlations with DCE-MRI radiomic features. We further developed prediction models for the expression enrichment scores of hallmark pathways based on radiomic features to expand the clinical application scenario of radiogenomics. Generally, our findings suggest that there was a significant correlation between the radiomic features and BC transcriptional activity and that this correlation can be validated and predicted in an external dataset. In addition, we believe that as more data are generated and combined with artificial intelligence technologies, such as deep learning, in the future, radiomics will certainly play an essential role in the clinical management of diseases throughout their stages, including early diagnosis, the monitoring of key molecular features, the selection of treatment strategies, and the evaluation of prognostic risk.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14010028/s1>, Figure S1: The inclusion and exclusion criteria of BC patients for radiogenomic cohorts. Figure S2: GSEA-based associations of different kinds of DCE-MRI features with hallmark pathways in the validation cohort of BC radiogenomics. Figure S3: Imaging features significantly associated with pathway activity were consistent to some extent in both cohorts. Figure S4: Quantitative associations of GSVA enrichment scores of hallmark pathways with imaging features in the discovery cohort. Figure S5: Quantitative associations of GSVA enrichment scores of hallmark pathways with imaging features in the validation cohort. Figure S6: The top 15 important radiomic features for the random forest prediction models of hallmark pathways. Table S1: The best-performed pre-processing configurations of RF models.

**Author Contributions:** Conceptualization, X.L., H.L. and X.S.; data curation, F.L., Y.Z., Y.B. and W.G.; formal analysis, W.M.; funding acquisition, Y.B., W.G., Y.L., X.L. and X.S.; investigation, Y.Z., Y.B. and W.G.; methodology, W.M., F.L., Y.Z. and H.L.; project administration, Y.L., X.L., H.L. and X.S.; resources, Y.Z., Y.B., W.G., Y.L. and X.L.; software, W.M. and F.L.; supervision, Y.L., X.L., H.L. and X.S.; validation, W.M. and Y.Z.; visualization, W.M. and F.L.; writing—original draft, W.M., Y.Z. and F.L.; writing—review and editing, W.M., H.L. and X.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 81830053, 61972084, and 61871121.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committees of Nanjing Medical University (protocol code 2019SR512 and date of approval 2019) for studies involving humans.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The RNA-seq data of the radiogenomics discovery cohort are available at <https://ngdc.cnbc.ac.cn/bioproject/> (GSA-Human number: HRA001100, accessed on 1 September 2019). The gene expression and MRI data of the validation cohort are available at <https://portal.gdc.cancer.gov/projects/TCGA-BRCA> and <https://wiki.cancerimagingarchive.net/display/Public/TCGA-BRCA> (accessed on 1 January 2020). Further information and other data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** We thank Wenjing Cui from the Department of Radiology, Jiangsu Province Hospital of Chinese Medicine for her contribution to tumor segmentation.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Polyak, K. Heterogeneity in breast cancer. *J. Clin. Investig.* **2011**, *121*, 3786–3788. [[CrossRef](#)] [[PubMed](#)]
3. Martelotto, L.G.; Ng, C.K.Y.; Piscuoglio, S.; Weigelt, B.; Reis-Filho, J.S. Breast cancer intra-tumor heterogeneity. *Breast Cancer Res.* **2014**, *16*, 210. [[CrossRef](#)] [[PubMed](#)]
4. Perou, C.M.; Sørlie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; Rees, C.A.; Pollack, J.R.; Ross, D.T.; Johnsen, H.; Akslén, L.A.; et al. Molecular portraits of human breast tumours. *Nature* **2000**, *406*, 747–752. [[CrossRef](#)] [[PubMed](#)]
5. Sørlie, T.; Perou, C.M.; Tibshirani, R.; Aas, T.; Geisler, S.; Johnsen, H.; Hastie, T.; Eisen, M.B.; van de Rijn, M.; Jeffrey, S.S.; et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 10869. [[CrossRef](#)] [[PubMed](#)]
6. Kennecke, H.; Yerushalmi, R.; Woods, R.; Cheang, M.C.U.; Voduc, D.; Speers, C.H.; Nielsen, T.O.; Gelmon, K. Metastatic Behavior of Breast Cancer Subtypes. *J. Clin. Oncol.* **2010**, *28*, 3271–3277. [[CrossRef](#)]
7. Waks, A.G.; Winer, E.P. Breast Cancer Treatment: A Review. *JAMA* **2019**, *321*, 288–300. [[CrossRef](#)]
8. Hammond, M.E.H.; Hayes, D.F.; Dowsett, M.; Allred, D.C.; Hagerty, K.L.; Badve, S.; Fitzgibbons, P.L.; Francis, G.; Goldstein, N.S.; Hayes, M.; et al. American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. *J. Clin. Oncol.* **2010**, *28*, 2784–2795. [[CrossRef](#)]
9. Dowsett, M.; Nielsen, T.O.; A'Hern, R.; Bartlett, J.; Coombes, R.C.; Cuzick, J.; Ellis, M.; Henry, N.L.; Hugh, J.C.; Lively, T.; et al. Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. *JNCI J. Natl. Cancer Inst.* **2011**, *103*, 1656–1664. [[CrossRef](#)]
10. Carey, L.A.; Perou, C.M.; Livasy, C.A.; Dressler, L.G.; Cowan, D.; Conway, K.; Karaca, G.; Troester, M.A.; Tse, C.K.; Edmiston, S.; et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *JAMA* **2006**, *295*, 2492–2502. [[CrossRef](#)]
11. Morra, A.; Jung, A.Y.; Behrens, S.; Keeman, R.; Ahearn, T.U.; Anton-Culver, H.; Arndt, V.; Augustinsson, A.; Auvinen, P.K.; Beane Freeman, L.E.; et al. Breast Cancer Risk Factors and Survival by Tumor Subtype: Pooled Analyses from the Breast Cancer Association Consortium. *Cancer Epidemiol. Biomark. Prev.* **2021**, *30*, 623–642. [[CrossRef](#)] [[PubMed](#)]
12. Gatzka, M.L.; Lucas, J.E.; Barry, W.T.; Kim, J.W.; Wang, Q.; Crawford, M.D.; Datto, M.B.; Kelley, M.; Mathey-Prevot, B.; Potti, A.; et al. A pathway-based classification of human breast cancer. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6994–6999. [[CrossRef](#)] [[PubMed](#)]
13. Kim, I.; Choi, S.; Kim, S. BRCA-Pathway: A structural integration and visualization system of TCGA breast cancer data on KEGG pathways. *BMC Bioinform.* **2018**, *19*, 42. [[CrossRef](#)] [[PubMed](#)]
14. Djomehri, S.I.; Gonzalez, M.E.; da Veiga Leprevost, F.; Tekula, S.R.; Chang, H.-Y.; White, M.J.; Cimino-Mathews, A.; Burman, B.; Basrur, V.; Argani, P.; et al. Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors. *Nat. Commun.* **2020**, *11*, 1723. [[CrossRef](#)]
15. Liberzon, A.; Birger, C.; Thorvaldsdóttir, H.; Ghandi, M.; Mesirov, J.P.; Tamayo, P. The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **2015**, *1*, 417–425. [[CrossRef](#)]
16. Lambin, P.; Leijenaar, R.T.H.; Deist, T.M.; Peerlings, J.; de Jong, E.E.C.; van Timmeren, J.; Sanduleanu, S.; Larue, R.T.H.M.; Even, A.J.G.; Jochems, A.; et al. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **2017**, *14*, 749–762. [[CrossRef](#)]
17. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2015**, *278*, 563–577. [[CrossRef](#)]
18. Pinker, K.; Chin, J.; Melsaether, A.N.; Morris, E.A.; Moy, L. Precision Medicine and Radiogenomics in Breast Cancer: New Approaches toward Diagnosis and Treatment. *Radiology* **2018**, *287*, 732–747. [[CrossRef](#)]
19. Grimm, L.J.; Mazurowski, M.A. Breast Cancer Radiogenomics: Current Status and Future Directions. *Acad. Radiol.* **2020**, *27*, 39–46. [[CrossRef](#)]
20. Bismeyer, T.; van der Velden, B.H.M.; Canisius, S.; Lips, E.H.; Loo, C.E.; Viergever, M.A.; Wesseling, J.; Gilhuijs, K.G.A.; Wessels, L.F.A. Radiogenomic Analysis of Breast Cancer by Linking MRI Phenotypes with Tumor Gene Expression. *Radiology* **2020**, *296*, 277–287. [[CrossRef](#)]
21. Fan, M.; Xia, P.; Clarke, R.; Wang, Y.; Li, L. Radiogenomic signatures reveal multiscale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nat. Commun.* **2020**, *11*, 4861. [[CrossRef](#)]
22. Ming, W.; Zhu, Y.; Bai, Y.; Gu, W.; Li, F.; Hu, Z.; Xia, T.; Dai, Z.; Yu, X.; Li, H.; et al. Radiogenomics analysis reveals the associations of dynamic contrast-enhanced-MRI features with gene expression characteristics, PAM50 subtypes, and prognosis of breast cancer. *Front. Oncol.* **2022**, *12*, 943326. [[CrossRef](#)]
23. Mazurowski, M.A.; Zhang, J.; Grimm, L.J.; Yoon, S.C.; Silber, J.I. Radiogenomic Analysis of Breast Cancer: Luminal B Molecular Subtype Is Associated with Enhancement Dynamics at MR Imaging. *Radiology* **2014**, *273*, 365–372. [[CrossRef](#)] [[PubMed](#)]

24. Xu, Z.; Wang, Y.; Chen, M.; Zhang, Q. Multi-region radiomics for artificially intelligent diagnosis of breast cancer using multimodal ultrasound. *Comput. Biol. Med.* **2022**, *149*, 105920. [[CrossRef](#)] [[PubMed](#)]
25. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [[CrossRef](#)]
26. Burnside, E.S.; Sickles, E.A.; Bassett, L.W.; Rubin, D.L.; Lee, C.H.; Ikeda, D.M.; Mendelson, E.B.; Wilcox, P.A.; Butler, P.F.; D’Orsi, C.J. The ACR BI-RADS® Experience: Learning From History. *J. Am. Coll. Radiol.* **2009**, *6*, 851–860. [[CrossRef](#)] [[PubMed](#)]
27. Xiao, J.; Rahbar, H.; Hippe, D.S.; Rendi, M.H.; Parker, E.U.; Shekar, N.; Hirano, M.; Cheung, K.J.; Partridge, S.C. Dynamic contrast-enhanced breast MRI features correlate with invasive breast cancer angiogenesis. *NPJ Breast Cancer* **2021**, *7*, 42. [[CrossRef](#)] [[PubMed](#)]
28. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 Bias Correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [[CrossRef](#)] [[PubMed](#)]
29. van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104. [[CrossRef](#)]
30. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
31. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
32. Anders, S.; Pyl, P.T.; Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **2015**, *31*, 166–169. [[CrossRef](#)]
33. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)] [[PubMed](#)]
34. Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545. [[CrossRef](#)]
35. Hänzelmann, S.; Castelo, R.; Guinney, J. GSEA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinform.* **2013**, *14*, 7. [[CrossRef](#)] [[PubMed](#)]
36. Pesapane, F.; Rotili, A.; Agazzi, G.M.; Botta, F.; Raimondi, S.; Penco, S.; Dominelli, V.; Cremonesi, M.; Jereczek-Fossa, B.A.; Carrafiello, G.; et al. Recent Radiomics Advancements in Breast Cancer: Lessons and Pitfalls for the Next Future. *Curr. Oncol.* **2021**, *28*, 2351–2372. [[CrossRef](#)] [[PubMed](#)]
37. Ferro, M.; de Cobelli, O.; Musi, G.; del Giudice, F.; Carrieri, G.; Busetto, G.M.; Falagarino, U.G.; Sciarra, A.; Maggi, M.; Crocetto, F.; et al. Radiomics in prostate cancer: An up-to-date review. *Ther. Adv. Urol.* **2022**, *14*, 17562872221109020. [[CrossRef](#)]
38. Qi, Y.; Zhao, T.; Han, M. The application of radiomics in predicting gene mutations in cancer. *Eur. Radiol.* **2022**, *32*, 4014–4024. [[CrossRef](#)]
39. Stanzione, A.; Verde, F.; Romeo, V.; Boccadifuoco, F.; Mainenti, P.P.; Maurea, S. Radiomics and machine learning applications in rectal cancer: Current update and future perspectives. *World J. Gastroenterol.* **2021**, *27*, 5306–5321. [[CrossRef](#)]
40. Cho, N. Breast Cancer Radiogenomics: Association of Enhancement Pattern at DCE MRI with Deregulation of mTOR Pathway. *Radiology* **2020**, *296*, 288–289. [[CrossRef](#)]
41. Yamamoto, S.; Han, W.; Kim, Y.; Du, L.; Jamshidi, N.; Huang, D.; Kim, J.H.; Kuo, M.D. Breast Cancer: Radiogenomic Biomarker Reveals Associations among Dynamic Contrast-enhanced MR Imaging, Long Noncoding RNA, and Metastasis. *Radiology* **2015**, *275*, 384–392. [[CrossRef](#)] [[PubMed](#)]
42. Bracken, A.P.; Ciro, M.; Cocito, A.; Helin, K. E2F target genes: Unraveling the biology. *Trends Biochem. Sci.* **2004**, *29*, 409–417. [[CrossRef](#)] [[PubMed](#)]
43. Hollern, D.P.; Swiatnicki, M.R.; Rennhack, J.P.; Misek, S.A.; Matson, B.C.; McAuliff, A.; Gallo, K.A.; Caron, K.M.; Andrechek, E.R. E2F1 Drives Breast Cancer Metastasis by Regulating the Target Gene FGF13 and Altering Cell Migration. *Sci. Rep.* **2019**, *9*, 10718. [[CrossRef](#)] [[PubMed](#)]
44. Oshi, M.; Takahashi, H.; Tokumaru, Y.; Yan, L.; Rashid, O.M.; Nagahashi, M.; Matsuyama, R.; Endo, I.; Takabe, K. The E2F Pathway Score as a Predictive Biomarker of Response to Neoadjuvant Therapy in ER+/HER2– Breast Cancer. *Cells* **2020**, *9*, 1643. [[CrossRef](#)] [[PubMed](#)]
45. Dong, Y.; Tu, R.; Liu, H.; Qing, G. Regulation of cancer cell metabolism: Oncogenic MYC in the driver’s seat. *Signal Transduct. Target. Ther.* **2020**, *5*, 124. [[CrossRef](#)] [[PubMed](#)]
46. Tang, M.; O’Grady, S.; Crown, J.; Duffy, M.J. MYC as a therapeutic target for the treatment of triple-negative breast cancer: Preclinical investigations with the novel MYC inhibitor, MYCi975. *Breast Cancer Res. Treat.* **2022**, *195*, 105–115. [[CrossRef](#)]
47. Kelsey, I.; Manning, B.D. mTORC1 Status Dictates Tumor Response to Targeted Therapeutics. *Sci. Signal.* **2013**, *6*, pe31. [[CrossRef](#)]
48. Faham, N.; Zhao, L.; Welm, A.L. mTORC1 is a key mediator of RON-dependent breast cancer metastasis with therapeutic potential. *NPJ Breast Cancer* **2018**, *4*, 36. [[CrossRef](#)]
49. Bao, Y.; Wang, L.; Shi, L.; Yun, F.; Liu, X.; Chen, Y.; Chen, C.; Ren, Y.; Jia, Y. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. *Cell. Mol. Biol. Lett.* **2019**, *24*, 38. [[CrossRef](#)]

50. Cox, T.R. The matrix in cancer. *Nat. Rev. Cancer* **2021**, *21*, 217–238. [[CrossRef](#)]
51. Jallow, F.; O’Leary, K.A.; Rugowski, D.E.; Guerrero, J.F.; Ponik, S.M.; Schuler, L.A. Dynamic interactions between the extracellular matrix and estrogen activity in progression of ER+ breast cancer. *Oncogene* **2019**, *38*, 6913–6925. [[CrossRef](#)] [[PubMed](#)]
52. Tecalco-Cruz, A.C.; Macías-Silva, M.; Ramírez-Jarquín, J.O.; Méndez-Ambrosio, B. Identification of genes modulated by interferon gamma in breast cancer cells. *Biochem. Biophys. Rep.* **2021**, *27*, 101053. [[CrossRef](#)] [[PubMed](#)]
53. Todorović-Raković, N.; Milovanović, J.; Greenman, J.; Radulovic, M. The prognostic significance of serum interferon-gamma (IFN- $\gamma$ ) in hormonally dependent breast cancer. *Cytokine* **2022**, *152*, 155836. [[CrossRef](#)] [[PubMed](#)]
54. Jia, Y.; Kodumudi, K.N.; Ramamoorthi, G.; Basu, A.; Snyder, C.; Wiener, D.; Pilon-Thomas, S.; Grover, P.; Zhang, H.; Greene, M.I.; et al. Th1 cytokine interferon gamma improves response in HER2 breast cancer by modulating the ubiquitin proteasomal pathway. *Mol. Ther.* **2021**, *29*, 1541–1556. [[CrossRef](#)]
55. Conti, A.; Duggento, A.; Indovina, I.; Guerrisi, M.; Toschi, N. Radiomics in breast cancer classification and prediction. *Semin. Cancer Biol.* **2021**, *72*, 238–250. [[CrossRef](#)]
56. Ming, W.; Li, F.; Zhu, Y.; Bai, Y.; Gu, W.; Liu, Y.; Sun, X.; Liu, X.; Liu, H. Predicting hormone receptors and PAM50 subtypes of breast cancer from multi-scale lesion images of DCE-MRI with transfer learning technique. *Comput. Biol. Med.* **2022**, *150*, 106147. [[CrossRef](#)]
57. Son, J.; Lee, S.E.; Kim, E.-K.; Kim, S. Prediction of breast cancer molecular subtypes using radiomics signatures of synthetic mammography from digital breast tomosynthesis. *Sci. Rep.* **2020**, *10*, 21566. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.