Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Mechanism-Driven Radiogenomic Analysis Reveals Potential Molecular Drivers of Pretreatment MRI Phenotypes in Breast Cancer

Tingfeng Zhang
Southern University of Science and Technology
Jinlin Yang
Macao Polytechnic University
Tingting Cui
South China University of Technology
Tao Tan
Macao Polytechnic University
Jiayue Xie
Jinan University
Huifen Zhang
Jinan University
Hong Hu
Southern University of Science and Technology
Gengxi Cai
First People's Hospital of Foshan
Jie Ma
Jinan University
Yipeng Yang
yang.yipeng@szhospital.com

Southern University of Science and Technology

Article

Keywords: Breast cancer, neoadjuvant chemotherapy, radiogenomics, residual cancer burden, deep learning

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11207582/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Mechanism-Driven Radiogenomic Analysis Reveals Potential Molecular Drivers of Pretreatment MRI Phenotypes in Breast Cancer

Tingfeng Zhang[1*], Jinlin Yang[3*], Tingting Cui[1,4], Tao Tan[3], Jiayue Xie[2], Huifen Zhang[2], Hong Hu[1], Gengxi Cai[5#]， Jie Ma[2#], Yipeng Yang[1#]

1.Division of Breast Surgery, Department of General Surgery, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen 518020, Guangdong, China
2.Department of Radiology, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, Shenzhen, China.
3.Faculty of Applied Sciences, Macao Polytechnic University, Macau, CO 999078, China
4.Key Laboratory of Biomedical Engineering of Guangdong Province, South China University of Technology, Guangzhou 510006, China.
5.Department of Breast Surgery, The First People's Hospital of Foshan (Foshan Hospital Affiliated to Southern University of Science and Technology), School of Medicine, Southern University of Science and Technology, Guangdong, China

Tingfeng Zhang and Jinlin Yang contribute equally to co-first author.

#Corresponding Author: Yipeng Yang. yang.yipeng@szhospital.com; Gengxi Cai: caigengxi2021@126.com; Jie Ma: cjr.majie@vip.163.com

***co-First Author:** Tingfeng Zhang: ztfmicle@163.com; Jinlin Yang: yangjinlin432@163.com

Other Author: Tingting Cui: tt1830156489tt@163.com; Tao Tan: Taotanjs@gmail.com; Jiayue Xie: 18842664032@163.com; Hong Hu: Chris_huhong@hotmail.com

Abstract

Purpose

To develop a pretreatment MRI model for predicting residual cancer burden (RCB) after neoadjuvant chemotherapy and a mechanism-driven radiogenomic framework to identify molecular programs underlying response-associated MRI phenotypes.

Methods

A multimodal deep learning model was developed using pretreatment MRI and clinical data from 492 patients across two centers. Radiomic phenotypes from 61 patients were integrated with public transcriptomic cohorts using differential expression, pathway enrichment, WGCNA, survival modeling, feature selection, and *in silico* perturbation analyses.

Results

The model achieved AUCs of 0.919 in training and 0.852 in testing. Of 195 response-associated differentially expressed genes, 109 showed concordant changes in an independent cohort. Response was linked to proliferation, DNA repair, hypoxia, immune-inflammatory, and estrogen-response programs. A 12-gene prognostic signature achieved a C-index of 0.798. Multilayer prioritization identified **ESR1, TBC1D9, GREB1**, and related genes, while **RARRES1** uniquely showed strong directional concordance between predicted and observed radiomic changes.

Conclusion

Pretreatment MRI heterogeneity reflects coordinated biological programs relevant to neoadjuvant treatment response. This framework supports noninvasive RCB stratification and generates testable mechanistic hypotheses for functional validation.

Keywords: Breast cancer; neoadjuvant chemotherapy; radiogenomics; residual cancer burden; deep learning

1.Introduce

Neoadjuvant chemotherapy (NAC) has become an important treatment strategy for locally advanced breast cancer and selected high-risk early-stage breast cancers. It can reduce tumor burden, increase the likelihood of breast-conserving surgery and complete surgical resection, and provide a unique window for evaluating in vivo treatment sensitivity, assessing early therapeutic response, and optimizing subsequent local and systemic treatment decisions[1, 2]. As key indicators for evaluating neoadjuvant treatment efficacy and predicting long-term outcomes, pathologic complete response (pCR) is generally associated with improved event-free survival and long-term prognosis[3]， whereas residual cancer burden (RCB) provides a quantitative assessment of residual disease and enables more refined risk stratification across different molecular subtypes[4]. However, both pCR and RCB require invasive pathological assessment. Therefore, identifying tumor aggressiveness, potential treatment sensitivity, and recurrence risk before the initiation of NAC remains a critical unmet need in precision breast cancer treatment[5].

Breast magnetic resonance imaging (MRI) is a valuable noninvasive tool for pretreatment assessment. At the macroscopic level, MRI captures spatial phenotypic information, including tumor morphology, enhancement patterns, signal distribution, margin characteristics, and intratumoral heterogeneity, and is an important imaging modality for evaluating local disease extent, treatment response, and residual disease in breast cancer[6]. . In recent years, radiomics and artificial intelligence have further enabled the quantitative analysis of MRI data, transforming medical images from visual records of anatomical structures into high-dimensional data that can characterize tumor phenotypes, treatment responses, and prognostic differences[7-9] . In the context of NAC for breast cancer, MRI-based radiomics and machine learning models have been used to predict pCR, molecular subtype, recurrence risk, and patient outcomes. These findings suggest that MRI phenotypes may, to some extent, reflect underlying biological differences in tumor cellular composition, angiogenesis, stromal architecture, necrosis and hypoxia, and the immune microenvironment[7, 10, 11]. Nevertheless, most existing studies have focused primarily on improving predictive performance, emphasizing metrics such as the area under the receiver operating characteristic curve (AUC), accuracy, and classification capability, while paying insufficient attention to the molecular basis and biological interpretability of imaging heterogeneity[12, 13] . In other words, MRI radiomics has demonstrated that imaging phenotypes can predict clinical outcomes; however, the molecular programs that collectively shape MRI-observed tumor heterogeneity, and how microscopic transcriptomic and pathway alterations are translated into macroscopically visible imaging phenotypes, remain incompletely understood.

Radiogenomics bridges noninvasive imaging phenotypes and tumor molecular biology by linking imaging features with gene expression, molecular subtypes, the immune microenvironment, and pathway activity[11, 14-16] . Previous breast cancer studies have suggested that MRI enhancement kinetics, imaging texture, and morphological features may be associated with molecular subtypes, recurrence risk, and gene-expression signatures[17] . Meanwhile, treatment response and tumor

evolution in breast cancer are not determined by a single gene but are jointly regulated by multiple biological programs, including estrogen receptor signaling, cell-cycle regulation, DNA damage repair, hypoxia adaptation, immune-inflammatory responses, angiogenesis, and epithelial–mesenchymal transition (EMT)[18] . Breast cancer radiogenomic research therefore needs to move beyond individual feature–gene correlations toward mechanistic interpretations at the levels of molecular networks and coordinated biological programs.

From a mechanistic perspective, these biological programs not only influence tumor-cell proliferation, invasion, and drug sensitivity but may also shape MRI-visible morphological, textural, and signal heterogeneity by altering cellular density, stromal composition, vascular permeability, necrosis and hypoxia, and immune infiltration[19, 20] Pretreatment MRI heterogeneity may therefore represent not an isolated imaging phenomenon, but rather the macroscopic projection of multiple coordinated molecular programs. Accordingly, biologically meaningful key genes should not merely be differentially expressed; they should exhibit concordant evidence across multiple dimensions, including gene-expression alterations, pathway activity, co-expression networks, prognostic relevance, and imaging-phenotype changes. Only within such a multilayered evidence framework can radiogenomic findings progress from correlational biomarkers toward mechanistic explanations of MRI phenotype formation.

Based on this rationale, we developed a mechanism-driven radiogenomic framework for breast cancer that integrates pretreatment MRI radiomic features, transcriptomic data, differential expression analysis, pathway activity analysis, weighted gene co-expression network analysis, and survival information. From a systems-biology perspective, this framework was designed to identify key molecular pathways and candidate driver genes associated with MRI-derived imaging heterogeneity. We further incorporated an approximate in silico gene perturbation strategy to infer potential changes in pathway activity following the perturbation of candidate genes. These pathway-level alterations were subsequently mapped, through a pathway–imaging feature correlation network, to predicted changes in imaging phenotypes. This approach established a continuous chain of evidence linking imaging phenotypes, biological programs, candidate genes, in silico perturbations, and concordant imaging changes. The proposed framework aims to elucidate how microscopic molecular alterations shape MRI-visible tumor phenotypes and to advance breast cancer radiogenomics from purely correlational analyses toward biologically interpretable mechanistic models.

2. Materials and Methods

2.1 Radiomics Cohorts and Characterization of MRI Heterogeneity

We included pretreatment MRI scans and corresponding clinical and prognostic data from 492 patients with breast cancer across two centers. The training cohort comprised 393 patients, including 316 and 77 patients from the two centers, respectively, while the internal validation cohort comprised 99 patients, including 79 and 20 patients from the respective centers. The dataset

included dynamic contrast-enhanced MRI (DCE-MRI), structured clinical variables—including age, TNM stage, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67 status—and matched RCB class data (Table 1). Patients with RCB-0/I were defined as responders, whereas those with RCB-II/III were defined as nonresponders.

For imaging preprocessing, all DCE-MRI scans from the 492 patients were resampled using linear interpolation to an isotropic voxel spacing of $2.0 \times 2.0 \times 2.0$ mm³ to minimize differences in spatial resolution arising from variations in scanners and imaging protocols. The resampled images were subsequently center-cropped or zero-padded to a uniform matrix size of $88 \times 176 \times 176$ voxels. This procedure ensured consistent network input dimensions while preserving as much of the central breast anatomy as possible. To reduce the influence of extremely high-intensity values on model training and improve intensity consistency across patients, image intensities were first clipped at the 99.5th percentile of the intensity distribution within each image volume. Min–max normalization was then applied to scale the intensity values to the range of [0, 1]. For the clinical data, missing values were encoded as -1 to indicate unavailable information.

Pretreatment breast MRI radiomic features and the corresponding clinical, survival, and pathological response data were obtained from a publicly available I-SPY–related cohort. After matching radiomic features with clinical data, 61 patients with available radiomic features and RCB class information were included in the subsequent mechanistic radiogenomic analysis (Table 2). Based on the RCB class determined after neoadjuvant treatment, patients were divided into low and high residual burden groups.

Radiomic features were first subjected to quality control. Features with high proportions of missing values or zero variance were removed, after which missing-value imputation and standardization were performed for the retained features. Ultimately, 65 radiomic features were included in the subsequent analyses. Using RCB-0/I versus RCB-II/III as the pathological residual burden endpoint, we evaluated differences in pretreatment MRI radiomic features between the two residual burden groups and identified candidate imaging features associated with RCB. The selected RCB-associated imaging features were then used to construct a radiomic heterogeneity score representing the residual burden–associated imaging heterogeneity phenotype captured by pretreatment MRI. This imaging heterogeneity phenotype was subsequently used for imaging–pathway–gene association analyses and concordance validation based on in silico perturbation.

2.2 Transcriptomic Cohorts and Data Preprocessing

To investigate the molecular basis underlying MRI heterogeneity, two publicly available transcriptomic cohorts of patients with breast cancer treated with neoadjuvant chemotherapy were included. GSE25066 was used as the discovery cohort, and GSE32603 served as an independent validation cohort. Raw gene-expression data were retrieved using the GEOquery package, and probes were mapped to gene symbols according to the corresponding platform annotation files. When multiple probes mapped to the same gene, their expression values were averaged to generate a gene-level expression matrix. The resulting expression matrices were subsequently normalized.

Based on the available treatment-response information, patients were classified as responders or nonresponders for subsequent differential expression analysis, pathway activity analysis, and co-expression network construction.

2.3 Systematic Characterization of Treatment Response–Associated Transcriptomic Features

To systematically characterize the molecular basis of response to neoadjuvant chemotherapy in breast cancer, treatment response–associated transcriptomic features were investigated at three levels: gene expression, pathway activity, and gene co-expression networks.

First, differential gene-expression analysis between responders and nonresponders was performed in the GSE25066 discovery cohort using the limma package. Differentially expressed genes (DEGs) were identified using the following criteria: an adjusted P value < 0.05 and an absolute log fold change (|logFC|) > 0.8. Given the differences in sample size, profiling platform, and clinical phenotype definitions between the cohorts, strict multiple-testing–adjusted significance was not simultaneously required in the independent GSE32603 validation cohort. Instead, DEGs identified in the discovery cohort were evaluated in the validation cohort by considering their detectability, concordance in the direction of expression change, and nominal statistical significance. A composite DEG score incorporating effect size and statistical significance was subsequently constructed, and candidate gene sets of varying sizes were generated for downstream systems-biology analyses.

Gene set variation analysis (GSVA) and single-sample gene set enrichment analysis (ssGSEA) were then performed to evaluate changes in pathway activity. Based on the Molecular Signatures Database (MSigDB) Hallmark gene sets, pathway activity scores were calculated for each sample in the GSE25066 and GSE32603 cohorts. Differences in pathway activity between responders and nonresponders were assessed using limma. For pathways that did not meet strict multiple-testing–adjusted significance in the validation cohort, concordance with the discovery cohort was further evaluated by considering the direction of change and nominal statistical significance. This strategy was used to identify key biological programs associated with treatment response.

To identify modules of coordinately expressed genes and characterize their potential functions, weighted gene co-expression network analysis (WGCNA) was performed in the GSE25066 cohort. To balance computational efficiency with the preservation of potentially relevant biological signals, highly variable genes were combined with the DEG gene set to construct a signed co-expression network. Network parameters were determined through soft-thresholding power analysis. Co-expression modules were subsequently identified using the dynamic tree-cutting algorithm, and a module eigengene was calculated for each module. Correlations of module eigengenes with treatment response–related phenotypes and pathway activity scores were then examined to identify functional modules closely associated with treatment sensitivity. The resulting module-level information served as an important source of evidence for the subsequent prioritization of key genes.

By integrating differential expression analysis, pathway activity analysis, and gene co-expression network analysis, we constructed a transcriptomic landscape of treatment response at the levels of individual genes, functional pathways, and coordinated regulatory networks. This landscape

provided the molecular foundation for subsequent radiogenomic association analyses and the identification of key candidate driver genes.

2.4 Construction of the Radiogenomic Association Map

To establish links between MRI phenotypes and their underlying molecular programs, radiogenomic correlation analyses were performed to systematically evaluate associations of imaging features with pathway activity, immune-related scores, and candidate gene expression. When radiomic and transcriptomic data could be matched at the individual-patient level, Spearman correlation analysis was used to calculate associations across paired samples. When direct patient-level matching was unavailable, group-level association analyses based on high versus low imaging heterogeneity were performed as a complementary approach. Pathway activity scores were derived from the GSVA results, whereas immune-related scores were calculated using predefined immune signature sets.

By constructing an "imaging feature–pathway activity–gene expression" association network, we established a continuous link between macroscopic MRI phenotypes and microscopic molecular characteristics, providing a basis for the subsequent identification of key driver genes and mechanistic inference.

2.5 Multilayer Evidence Integration and Identification of Candidate Driver Genes

To identify key molecular factors that may drive the formation of MRI phenotypes, we developed a multilayer evidence-integration framework to systematically prioritize candidate genes. Candidate genes were derived from multiple sources of evidence, including differential expression analysis, independent-cohort validation, co-expression network analysis, pathway activity analysis, radiogenomic correlation analysis, and prognostic analysis. In samples with complete follow-up information, univariable Cox proportional hazards models, random survival forests (RSFs), and LASSO–Cox models were used to evaluate associations between gene expression and patient outcomes, thereby further identifying candidate genes with potential clinical relevance.

Based on these results, a composite scoring system was constructed that incorporated multiple dimensions of evidence, including differential expression, concordance in the validation cohort, co-expression network characteristics, pathway associations, radiogenomic associations, survival relevance, model-based selection, functional annotations, in silico perturbation effects, and imaging concordance. All component scores were normalized to a range of 0–1 and combined using prespecified weights to calculate a Final Priority Score.

According to their composite scores, candidate genes were classified as high-priority mechanistic drivers, intermediate-priority candidates, or low-priority candidates. To qualify as a high-priority mechanistic driver, a gene was required to have a high composite score, demonstrate favorable directional concordance across predictive analyses, and be supported by differential expression or co-expression network analysis.

By integrating transcriptomic features, prognostic information, and radiogenomic evidence, this framework aimed to identify key candidate driver genes linking treatment response–associated molecular programs to the formation of MRI phenotypes.

2.6 Mechanism-Driven In Silico Perturbation Framework Based on the "Gene–Pathway–Imaging" Cascade

To further infer the potential effects of candidate genes on the formation of MRI phenotypes, we developed a mechanism-driven in silico gene perturbation framework based on bulk RNA sequencing data. By establishing a continuous mapping from genes to pathway activity and subsequently to imaging features, this framework was designed to approximate the biological effects that may result from gene perturbation.

First, Spearman correlation coefficients were calculated between candidate gene-expression levels and GSVA-derived pathway activity scores. These correlations were then used to infer potential changes in pathway activity following the in silico inhibition of each gene. Next, using the associations between pathway activity and imaging features, the predicted pathway-level changes were mapped to the MRI feature level, generating a gene-specific Predicted Radiomic Shift.

The direction of each predicted radiomic shift was subsequently compared with the observed differences in MRI features between the high- and low-imaging-heterogeneity groups. The Consistency Ratio and Weighted Consistency Score were used to quantify concordance between the predicted and observed imaging changes.

Through this framework, we sought to reconstruct a continuous biological cascade linking alterations in gene activity to pathway remodeling and, ultimately, changes in MRI phenotypes. This approach was intended to advance radiogenomic research from conventional correlation-based analyses toward models capable of providing mechanistic interpretations.

2.8 Development of an Imaging- and Clinical Data–Based Model for RCB Classification

To accurately classify RCB following neoadjuvant treatment for breast cancer, we developed a multimodal deep learning framework integrating DCE-MRI with structured clinical information. The proposed model used only MRI data and patient clinical information, without incorporating additional modalities or handcrafted features, thereby ensuring consistency of model inputs and facilitating clinical interpretability.

DCE-MRI Feature Extraction: For three-dimensional breast MRI data, the images were first input into a Mamba-based 3D visual encoder designed to learn global long-range dependencies and local spatial structural features. The encoder comprised four hierarchical stages with progressive downsampling. At each stage, volumetric data were successively abstracted through a multistage feature-extraction module, yielding multiscale features $I=[I_1, I_2, I_3, I_4]$, with progressively increasing dimensions of [64, 128, 256, 512] . During multiscale feature extraction, the output features from each stage were enhanced using a global state-space modeling module to improve the representation of tumor heterogeneity and complex boundary structures. The highest-level features were subsequently processed using a three-dimensional UNETR BasicBlock to further integrate local contextual information and generate a deep imaging representation. Finally, 3D adaptive average pooling was applied to globally aggregate the voxel-level features into a fixed-dimensional imaging feature vector for subsequent multimodal fusion and classification.

Encoding of Structured Clinical Information: A Clinical Transformer based on multilayer feature alignment was designed to encode structured clinical variables, including age, TNM stage, ER, PR, HER2, and Ki-67 status. Through a hierarchical mapping mechanism, the encoder transformed the original low-dimensional clinical variables into multiscale representations aligned with the corresponding imaging feature hierarchies, denoted as$C=[C_1, C_2, C_3, C_4]$, with dimensions of [64, 128, 256, 512] . This design enabled the alignment and interactive modeling of clinical and imaging features across different semantic levels.

Imaging–Clinical Feature Alignment and Expert Modeling: To enable efficient interactions between imaging features and clinical information, a dual-branch representation modeling strategy based on Expert Learning was introduced during multimodal fusion. Specifically, the model first extracted global imaging features (I_i) and structured clinical features (C_i) separately and projected them into a shared 512-dimensional feature space to ensure consistency between the cross-modal representations. An imaging expert branch and a clinical expert branch were then constructed to perform independent nonlinear modeling of the respective modality-specific features:

$$E_{img} = \phi_{img}(I), \quad E_{clin} = \phi_{clin}(C) \tag{1}$$

where ($\phi_{img}(\cdot)$) and ($\phi_{clin}(\cdot)$) are each composed of a linear layer followed by a ReLU activation function to learn modality-specific representations. This design avoids representational conflicts caused by the premature fusion of heterogeneous modalities, thereby improving the stability and discriminative capacity of the learned features.

Adaptive Imaging–Clinical Fusion: After obtaining the imaging and clinical expert features, we further introduced a gated expert-fusion module to enable dynamic, data-driven, weighted cross-modal fusion. Specifically, the imaging and clinical features were concatenated and then fed into a gating network to learn adaptive fusion weights:

$$G = \sigma(W_g[I, C] + b_g) \tag{2}$$

where ([I, C]) denotes the concatenated representation of the imaging and clinical features,($\sigma(\cdot)$) denotes the sigmoid function, and ($G \in [0, 1]^d$) represents the element-wise gating weight vector. The outputs of the two expert branches were then adaptively fused through the gating mechanism:

$$F_{fusion} = G \odot E_{img} + (1 - G) \odot E_{clin} \tag{3}$$

where ($\odot$) denotes element-wise multiplication. This mechanism adaptively adjusts the contribution of each modality according to the degree of imaging–clinical concordance within each sample, thereby reducing the information bias associated with fixed-weight fusion strategies. Finally, the fused feature representation (F_{fusion}) was fed into a multilayer perceptron (MLP) classifier to generate the predicted RCB class.

$$\hat{y} = \mathrm{MLP}\left(F_{fusion}\right) \tag{4}$$

2.7 Statistical Analysis

All statistical analyses were performed using R software (version 4.3.2). Unless otherwise specified, continuous variables were compared between groups using the Wilcoxon rank-sum test, and correlations were assessed using Spearman correlation analysis. The Benjamini–Hochberg procedure was applied to control the false discovery rate (FDR) for multiple comparisons.

Differential expression and pathway activity analyses were performed using the limma package. GSVA and ssGSEA were used to assess pathway activity. WGCNA was conducted using a signed network. Survival analyses were performed using univariable Cox proportional hazards models, RSFs, and LASSO–Cox regression models. The performance of the prognostic models was evaluated using Kaplan–Meier survival curves, time-dependent receiver operating characteristic curves, and the concordance index (C-index). Unless otherwise specified, a two-sided P value < 0.05 was considered statistically significant.

3. Results

3.1 Association Between Pretreatment MRI Radiomic Phenotypes and Residual Cancer Burden

This study included 492 patients with breast cancer from two centers, all of whom had pretreatment DCE-MRI scans, clinical data, and matched RCB class information. Of these patients, 393 were included in the training cohort, comprising 316 patients from Center 1 and 77 from Center 2. The remaining 99 patients were included in the internal validation cohort, comprising 79 patients from Center 1 and 20 from Center 2. The structured clinical variables primarily included age, TNM stage, ER, PR, HER2 and Ki-67 status. The baseline characteristics, sample composition, and specific roles of the individual cohorts in this study are summarized in Table 1.

A multimodal DCE-MRI–based model for RCB prediction was developed using data from these 492 patients. The model demonstrated relatively stable discriminative performance in both the training and internal validation cohorts. In the overall training cohort, the model achieved an AUC of 0.919, with center-specific AUCs of 0.935 and 0.899 for Centers 1 and 2, respectively. Despite potential intercenter heterogeneity in patient composition, MRI scanners, and imaging protocols, the model maintained predictive performance in both centers. In the overall internal validation cohort, the AUC was 0.852, representing only a modest decrease from the training cohort and indicating favorable internal generalizability. Center-specific evaluation yielded internal validation AUCs of 0.867 for Center 1 and 0.860 for Center 2 (Figure 1).

Further center-stratified analyses showed that the model maintained a certain degree of predictive performance across data from different centers, suggesting robustness to multicenter DCE-MRI data. Notably, despite the relatively small sample size at Center 2, the model achieved a high AUC in its internal validation cohort, indicating that its predictive performance was not exclusively dependent on data from the center with the larger sample size. In addition, spatial normalization and intensity

standardization were applied during image preprocessing to minimize nonbiological variation introduced by differences in MRI scanners and imaging protocols, thereby improving the consistency of multicenter MRI data and the stability of model predictions.

To further investigate the potential biological relationship between pretreatment MRI phenotypes and response to neoadjuvant treatment, we integrated publicly available I-SPY radiomic data with clinical phenotype information from GSE25066. Patient-level matching identified 61 patients with available pretreatment MRI radiomic features, clinical data, and RCB class information. Their detailed clinical characteristics stratified by RCB class are presented in Table 2. Following quality control based on missingness, zero variance, and data completeness, 65 quantitative MRI radiomic features were retained for subsequent analyses.

Using RCB class as the clinical response anchor, patients were divided into RCB-0/I and RCB-II/III groups. The RCB-0/I group comprised 14 patients and represented a lower residual tumor burden and a more favorable treatment response, whereas the RCB-II/III group comprised 47 patients and represented a higher residual tumor burden and a relatively unfavorable treatment response (Figure 2A). Differential analysis of pretreatment MRI radiomic features was subsequently performed between the two groups. The volcano plot showed that several morphological, first-order statistical, and texture features exhibited relatively large, standardized effect sizes and nominally significant trends. However, after adjustment for multiple testing, no individual radiomic feature reached the FDR significance threshold (Figure 2B).

After integrated ranking based on effect size, statistical significance, and feature stability, the highest-ranked radiomic features primarily reflected first-order signal-intensity dispersion, gray-level run-length distributions, and gray-level dependence structures. The complete ranking of RCB-associated radiomic features is provided in Supplementary Table S1. The heatmap demonstrated coordinated differences in these highly ranked features between patients with RCB-0/I and those with RCB-II/III, while also revealing substantial interpatient heterogeneity (Figure 2C).

Given that a single radiomic feature may be insufficient to capture the complex spatial heterogeneity of tumors, we integrated the 20 highest-ranked RCB-associated radiomic features to construct a continuous radiomic heterogeneity score. When patients were ordered according to this score, the score distributions showed a degree of separation between the RCB-0/I and RCB-II/III groups. The radiomic heterogeneity score was also nominally associated with RCB group membership (Wilcoxon $P = 0.0439$; Figure 2D).

These findings suggest that the pretreatment MRI representation of residual tumor burden following neoadjuvant treatment may not be driven by an isolated imaging metric. Instead, it may reflect a composite imaging phenotype jointly determined by multiple dimensions, including tumor morphology, textural architecture, and signal-intensity distribution. Given the relatively limited sample size of the matched radiomics cohort, these findings are more appropriately considered an imaging-phenotype foundation for subsequent radiogenomic analyses than definitive evidence supporting any individual imaging biomarker.

3.2 Treatment Response–Associated Transcriptomic Features Reveal Distinct Molecular Programs

To investigate the molecular basis underlying MRI phenotypes, GSE25066 was used as the discovery cohort to compare gene-expression profiles between responders and nonresponders. Tiered threshold-based screening ultimately identified 195 differentially expressed genes in the discovery cohort (discovery DEGs). The complete differential expression results for both transcriptomic cohorts are provided in Supplementary Table S2. The volcano plot demonstrated clear differences in the direction and statistical significance of treatment response–associated gene-expression changes (Figure 3A). The heatmap of the top 50 DEGs further illustrated distinct expression patterns between responder/pCR and nonresponder/residual disease (RD) samples (Figure 3C).

Multiple genes involved in cell-cycle regulation and DNA replication, including **MCM2, MCM3, MCM6, MELK**, and **CDKN2A**, were upregulated in responders. In contrast, nonresponders exhibited higher expression of genes associated with luminal differentiation and endocrine signaling, including **ESR1** and **GLI3**. These findings suggest that tumors with a more favorable treatment response tend to exhibit active proliferative programs, whereas nonresponsive tumors may retain stronger hormone receptor–related characteristics (Figure 3A and C).

GSE32603 was used as an independent validation cohort. Because no gene reached statistical significance after FDR correction in this cohort, validation was not based on a strict intersection of FDR-significant genes across the two cohorts. Instead, gene detectability on the validation platform, concordance in the direction of effect, and nominal P values were considered collectively. The volcano plot for the validation cohort was therefore used to illustrate overall effect trends rather than to identify an independent set of FDR-significant DEGs (Figure 3B).

Among the 195 discovery DEGs, 142 were detectable in GSE32603, of which 109 showed concordant directions of expression change across the two cohorts. Among these directionally concordant genes, 30 were supported by a nominal P value < 0.05 in the validation cohort, while 42 were supported by a nominal P value < 0.10. The scatterplot comparing log fold changes (logFCs) between the discovery and validation cohorts illustrates the degree of directional concordance and nominal statistical support (Figure 3D; Supplementary Table S3).

Overall, differential expression analysis identified two relatively distinct treatment response–associated transcriptional states: a proliferation/cell-cycle–related state and a luminal/endocrine differentiation state. Although no genes reached FDR-adjusted significance in the validation cohort, the observed directional concordance supports a degree of cross-cohort stability in the discovery signals and provides a candidate gene basis for subsequent integrative analyses at the pathway and network levels.

3.3 Coordinated Pathway-Level Molecular Programs Are Associated With Response to Neoadjuvant Treatment

To characterize treatment response–associated transcriptomic alterations at the functional level, GSVA/ssGSEA was used to calculate sample-level activity scores for the MSigDB Hallmark pathways. The heatmap of key pathways in GSE25066 revealed distinct pathway activity patterns between responders and nonresponders across programs related to the cell cycle, DNA repair, hypoxia, immune-inflammatory activity, and estrogen response (Figure 4A).

Differential pathway analysis showed that **MYC Targets V1, E2F Targets, G2M Checkpoint, DNA Repair, Hypoxia, Interferon-gamma Response, TNF/NF-kappaB Signaling**, and **mTORC1 Signaling** exhibited higher activity in responders. In contrast, **Estrogen Response Early** and **TGF-beta Signaling** showed higher activity in nonresponders (Figure 4B). Boxplots further illustrated the between-group distributions of representative pathways, including **Estrogen Response Early, G2M Checkpoint, Hypoxia**, and **Inflammatory Response** (Figure 4D).

Although no pathway reached statistical significance after FDR correction in GSE32603, key programs—including **Estrogen Response, TNF/NF-kappaB Signaling, Interferon-gamma Response, G2M Checkpoint, Inflammatory Response, E2F Targets, MYC Targets**, and **Hypoxia**—continued to exhibit biologically concordant trends. Comparison of pathway-level logFCs between the discovery and validation cohorts demonstrated concordant directions of effect for 10 prespecified key pathways (Figure 4C; Supplementary Table S4).

These findings indicate that treatment response is not determined by isolated genes but is jointly shaped by multiple coordinated biological programs involving proliferation, DNA repair, hypoxia adaptation, immune-inflammatory responses, and hormonal signaling.

3.4 WGCNA Identifies Co-expression Modules Linking Treatment Response, RCB, and Pathway Activity

In the GSE25066 discovery cohort, highly variable genes and the DEG_top200 candidate genes were used to construct a signed WGCNA network. Soft-thresholding analysis jointly evaluated scale-free topology fit and mean connectivity, and a soft-thresholding power of $\beta = 3$ was selected for network construction (Figure 5A). Nine co-expression modules were ultimately identified: turquoise, blue, brown, yellow, green, red, black, pink, and magenta. Complete module gene assignments and module–trait correlation results are provided in Supplementary Tables S5A–B. The module eigengene clustering dendrogram illustrated the correlation structure among the modules (Figure 5B).

The module–trait correlation heatmap showed that multiple modules were significantly associated with treatment response, pCR, RCB-related variables, and Hallmark pathway activity (Figure 5C). The black module was positively correlated with pCR/treatment response and strongly associated with **G2M Checkpoint, E2F Targets, MYC Targets**, and **DNA Repair**, indicating a proliferative and DNA damage-repair program. The magenta module was negatively correlated with pCR/treatment response and positively correlated with RCB class, suggesting a potential association with treatment resistance and greater residual disease burden.

The yellow and pink modules were closely associated with estrogen-response programs. The brown module was strongly associated with **EMT, Angiogenesis, TGF-beta Signaling**, and **Notch Signaling**, whereas the red and turquoise modules were primarily associated with **Interferon-gamma Response, Inflammatory Response**, and **TNF/NF-kappaB Signaling**.

A simplified network diagram illustrates the key modules, representative intramodular hub genes, and their principal phenotypic and pathway associations (Figure 5D). These modules provide an intermediate organizational layer between single-gene differential expression and pathway activity. Candidate genes could therefore be evaluated according to whether they were embedded within coordinated expression networks associated with treatment response, RCB, and key biological pathways, thereby strengthening the biological credibility of subsequent mechanistic candidate-gene prioritization.

3.5 Imaging–Molecular Associations and Multilayer Evidence Integration Identify Prognostically Relevant Candidate Hub Genes

To link pretreatment MRI phenotypes with potential molecular programs, we first examined Spearman correlations between selected radiomic features and GSVA-derived pathway activity scores. Shape-related features exhibited the strongest pathway associations. Specifically, original_shape_Maximum2DDiameterColumn and original_shape_LeastAxisLength were nominally positively correlated with **Hypoxia** activity（r = 0.428，P = 5.72 x 10^-4；r = 0.420，P = 7.57 x 10^-4, respectively）In contrast, the radiomic heterogeneity score was negatively correlated with **TNF/NF-kappaB Signaling**（r = -0.352，P = 0.00535；Figure 6A）.At the gene level, first-order intensity-dispersion features were associated with endocrine-related genes, including **SLC39A6, FOXA1, ESR1, XBP1**, and **DNAJC12**, whereas shape features were associated with candidate genes such as **RARRES1** and **SCUBE2** (Figure 6B). None of these imaging–pathway or imaging–gene correlations remained significant after FDR correction. They were therefore considered exploratory evidence for constructing the subsequent mechanistic inference network rather than definitive evidence supporting independent biomarkers.

Evidence from differential expression, directional concordance in the validation cohort, WGCNA module membership, pathway associations, radiogenomic correlations, survival associations, RSF/LASSO–Cox model selection, functional annotations, in silico perturbation magnitude, and directional imaging concordance was subsequently integrated to prioritize 8,966 candidate genes. The complete candidate gene pool and corresponding source labels are provided in Supplementary Table S7. The highest-ranked genes according to the composite score included **ESR1, TBC1D9, GREB1, AGR2, SLC39A6, SCUBE2, FOXA1, TTK, MLPH**, and **GATA3** (Figure 6C–D; Supplementary Table S8). These candidates were primarily involved in luminal differentiation and estrogen signaling, cell-cycle and mitotic regulation, and prognostically relevant molecular networks.

Among 508 patients with available gene-expression profiles and distant recurrence-free survival (DRFS) data, univariable Cox analysis identified 75 nominally significant prognostic genes, of

which 72 remained significant after FDR correction. LASSO–Cox analysis ultimately retained 12 genes for constructing the prognostic risk score. Complete results from the univariable Cox, RSF, and LASSO–Cox analyses are provided in Supplementary Tables S6A–C.

Using the median risk score as the cutoff, patients were divided into high- and low-risk groups, each comprising 254 patients. There were 85 DRFS events in the high-risk group and 26 in the low-risk group. The high-risk group had significantly poorer DRFS than the low-risk group（log-rank P = 6.13 x 10^-12；Figure 6E）. The time-dependent AUCs of the risk score for predicting 3- and 5-year DRFS were 0.795 and 0.742, respectively (Figure 6F). A nomogram integrating the risk score with clinical variables achieved a C-index of 0.7982 among 485 complete cases with 106 DRFS events (Supplementary Table S6D).

Notably, a high overall composite score did not necessarily indicate that perturbation of a candidate gene could reproduce the actual changes in MRI heterogeneity. Accordingly, genes such as **ESR1, TBC1D9**, and **GREB1** were considered important candidate regulators supported by extensive multi-omics evidence. However, their classification as mechanistic driver genes required further evaluation based on the directional concordance between predicted imaging changes following in silico perturbation and the observed MRI heterogeneity patterns (Figure 7D–F).

3.6 In Silico Gene Perturbation Predicts MRI Phenotypic Changes Through Pathway Remodeling and Evaluates Their Directional Concordance

To further assess whether candidate genes may contribute to the formation of MRI phenotypes, we constructed a continuous Gene–Pathway–MRI mapping framework. This analysis did not involve experimental gene editing. Instead, correlations between candidate gene expression and GSVA-derived pathway activity in bulk RNA-seq data were used to approximate the pathway alterations that might occur following in silico gene inhibition. This analysis ultimately generated 179,320 predicted gene–pathway perturbation relationships and 188,286 gene-level predicted radiomic changes.

Different candidate genes exhibited distinct predicted patterns of pathway remodeling (Figure 7A–B). In silico inhibition of **RARRES1** was predicted to enhance **Estrogen Response Early** while suppressing **Hypoxia, Inflammatory Response, TNF/NF-kappaB Signaling, Interferon-gamma Response, Angiogenesis, MYC Targets**, and **Notch Signaling**. Its strongest predicted effects included activation of **Estrogen Response Early** (predicted delta = 0.608) and suppression of **Hypoxia** (predicted delta = −0.484), **Inflammatory Response** (predicted delta = −0.462), and **TNF/NF-kappaB Signaling** (predicted delta = −0.454).

In contrast, in silico inhibition of **ESR1** produced an opposing pattern of predicted pathway changes, characterized by decreased **Estrogen Response Early** activity and increased activity of **MYC Targets, Hypoxia, E2F Targets, G2M Checkpoint**, and immune-inflammatory programs. Following the in silico inhibition of the proliferation-related genes **TTK** and **MELK**, the activities of **G2M Checkpoint, E2F Targets, MYC Targets, DNA Repair, Hypoxia**, and **mTORC1 Signaling** were predicted to decrease, whereas **Estrogen Response** and **KRAS Signaling Up** were

predicted to increase. The complete predicted pathway changes following in silico inhibition of all candidate genes are provided in Supplementary Table S9.

The predicted pathway alterations were subsequently propagated to the imaging level through the pathway–radiomic correlation network, generating predicted MRI feature changes for each candidate gene (Figure 7C). Inhibition of **RARRES1** was predicted to decrease shape- and intensity-dispersion features, including maximum 2D diameter, least-axis length, minor-axis length, root mean squared intensity, variance, and mean absolute deviation. Inhibition of **ESR1** was predicted to increase multiple shape and intensity-dispersion features while decreasing gray-level size-zone matrix (GLSZM) Zone Entropy.

Inhibition of **TTK** and **MELK** was primarily predicted to decrease proliferation-related texture features, including Size-zone Nonuniformity and Small-area Emphasis, while increasing Zone Entropy and several first-order intensity features. The complete gene-level predicted radiomic changes are presented in Supplementary Table S10A.

To assess the biological plausibility of the predicted radiomic changes, their directions following in silico perturbation were compared with the observed radiomic differences between the high- and low-MRI-heterogeneity groups (Figure 7D). In the high-heterogeneity group, six radiomic features were increased and 14 were decreased. Among all 8,966 candidate genes, 2,824 achieved a Consistency Ratio of at least 0.70. The Consistency Ratios of representative candidate genes and the prespecified cutoff of 0.70 are shown in Figure 7E. The Weighted Consistency Score incorporated the magnitudes of both observed and predicted radiomic changes and was used to distinguish weighted directional concordance from directional conflict (Figure 7F). Directional imaging-concordance statistics for all candidate genes are provided in Supplementary Table S10B. **RARRES1** achieved a Consistency Ratio of 0.70 and a Weighted Consistency Score of 0.995, indicating strong directional concordance between its predicted radiomic changes and the observed MRI heterogeneity pattern. **TTK** and **MELK** both had Consistency Ratios of 0.55, with Weighted Consistency Scores of 0.900 and 0.899, respectively. In contrast, although **ESR1, TBC1D9**, and **GREB1** had high overall priority scores, their Weighted Consistency Scores were negative, indicating that the predicted directions of their perturbation effects did not directly reproduce the observed MRI differences between the high- and low-heterogeneity groups. After jointly considering the final priority ranking, Consistency Ratio, Weighted Consistency Score, and supporting evidence from DEG and WGCNA analyses, **RARRES1** was the only high-priority candidate mechanistic driver that met all prespecified stringent criteria (Figure 7D–F).

Overall, this study established a continuous chain of evidence linking MRI phenotypes, radiomic heterogeneity, pathway activity alterations, candidate hub genes, in silico gene perturbations, and predicted radiomic phenotype shifts. The findings support the hypothesis that pretreatment MRI heterogeneity is jointly shaped by multiple coordinated molecular programs, including estrogen response, cell-cycle regulation, DNA repair, hypoxia adaptation, immune-inflammatory activity, and EMT. Under the stringent directional-concordance criteria, **RARRES1** exhibited the strongest

evidence as a mechanistic driver, whereas **ESR1, TBC1D9**, and **GREB1** represented important candidate regulators supported by multiple layers of evidence.

4. Discussion

This study demonstrates that heterogeneity on pretreatment breast MRI does not arise solely from differences in tumor size or morphology; rather, it may represent the macroscopic imaging manifestation of multiple coordinated biological processes. The high-heterogeneity phenotype was characterized by increased texture granularity and regional nonuniformity, accompanied by decreases in several shape and intensity-dispersion features. These findings suggest that MRI captures more than a simple tumor-volume effect and may reflect the remodeling of intratumoral architecture, the spatial organization of tumor cells, and tissue composition. Previous breast MRI studies have similarly shown that pretreatment enhancement patterns, perfusion heterogeneity, and dynamic imaging characteristics are closely associated with response to neoadjuvant treatment and tumor heterogeneity.[21, 22]. From a broader radiomics perspective, medical images can capture differences in tumor phenotypes through high-dimensional quantitative features and may, to some extent, encode information related to the tumor microenvironment, spatial architecture, and prognosis[23, 24]. In this study, transcriptomic and pathway analyses further suggested that MRI phenotypes may simultaneously encode multiple biological states, including cellular proliferation, DNA damage repair, hypoxia adaptation, immune-inflammatory responses, EMT and estrogen signaling. MRI heterogeneity is therefore more likely to represent a system-level, cross-scale phenotype than a feature determined by any single gene or pathway. These findings support the view that MRI provides not only anatomical information but also a potential noninvasive window into the molecular state of tumors and the complexity of their microenvironment.

Multilayer evidence integration further indicated that multiple biological programs collectively constitute the molecular basis of MRI phenotype formation. Genes including **ESR1, FOXA1, GATA3, GREB1, AGR2**, and **SLC39A6** collectively implicate the luminal differentiation and estrogen receptor signaling axis, which plays a central role in the molecular classification, endocrine dependence, and treatment response of breast cancer[25-28]. In contrast, candidate genes such as **TTK, MELK, CEP55**, and **NDC80** predominantly reflect cell-cycle progression, chromosome segregation, and proliferative activity. This finding is consistent with the enrichment of **MYC Targets, E2F Targets, G2M Checkpoint**, and **DNA Repair** pathways in the treatment response–related analyses[29-31] . Meanwhile, modules associated with hypoxia, EMT, angiogenesis, and immune-inflammatory activity suggest that remodeling of the tumor microenvironment is another important source of MRI heterogeneity. By influencing cellular density, stromal composition, vascular permeability, necrotic and hypoxic regions, and immune infiltration, these programs may further alter MRI-visible texture, morphology, and signal distributions[19].

Notably, the genes with the highest overall priority scores were not necessarily the most likely mechanistic drivers. **ESR1, TBC1D9**, and **GREB1** received strong support from differential expression, co-expression network, survival, and machine learning analyses, indicating that they are

important radiogenomic candidate biomarkers. However, **RARRES1** was the only gene that simultaneously met the criteria for multilayer evidence support and directional imaging concordance following in silico perturbation. This distinction is important: genes supported by extensive associative evidence may represent disease subtype, prognostic state, or the broader biological context of treatment sensitivity, whereas a mechanistic driver should additionally satisfy the requirement that its perturbation influences relevant pathways through a biologically plausible route and reproduces the observed direction of imaging-phenotype changes.In this study, **RARRES1** demonstrated strong directional concordance between its predicted and observed radiomic changes, suggesting that it may occupy a key position linking molecular-state alterations, pathway remodeling, and MRI phenotype formation. This finding does not establish **RARRES1** as a causal driver. Rather, it identifies **RARRES1** as a high-priority candidate for subsequent validation using cellular experiments, organoid models, and animal imaging studies.

Conventional radiogenomic studies have largely focused on statistical associations between individual imaging features and the expression of single genes. Such analyses may not adequately capture the coordinated networks of pathways, co-expression modules, and microenvironmental programs underlying tumor imaging phenotypes. They also have limited ability to distinguish genes that are merely correlative biomarkers from those that may actively contribute to the formation of imaging heterogeneity[32] . We therefore sought to establish a continuous chain of evidence linking gene-expression alterations, pathway activity remodeling, and imaging-feature changes. By integrating GSVA, WGCNA, prognostic modeling, and approximate in silico perturbation inference, we extended statistical associations into directional mechanistic hypotheses. It should be emphasized that this framework cannot establish causality in the strict sense. Nevertheless, it can prioritize candidate genes whose predicted effects may propagate through biologically plausible pathways and partially reproduce the observed patterns of MRI change. The principal innovation of this study therefore lies not in proposing a new predictive model, but in extending radiomics from descriptive association analysis toward mechanism-oriented radiogenomic research.

From a clinical perspective, pretreatment MRI heterogeneity patterns may provide noninvasive information about the molecular state of breast cancer and serve as an important complement to tissue biopsy. Although tissue biopsy provides high-resolution molecular information from a localized region, it may not comprehensively capture the spatial heterogeneity of the entire tumor. In contrast, MRI can characterize differences in tumor morphology, perfusion, enhancement, and texture at the whole-tumor level[33] . It may therefore be particularly suitable for assessing spatially heterogeneous biological states such as hypoxia, immune-inflammatory activity, endocrine differentiation, and microenvironmental remodeling. Integrating radiomics with transcriptomics, pathway analysis, and prognostic information may further improve patient stratification before NAC and facilitate the identification of potential therapeutic targets. Of particular interest, a mechanism-driven radiogenomic framework can not only identify candidate genes associated with treatment response or prognosis but also generate biologically interpretable, directional hypotheses, thereby

providing a theoretical basis for subsequent functional experiments and the exploration of therapeutic strategies.

This study has several limitations. First, although the model demonstrated a degree of stability across multicenter data, differences in MRI scanners, imaging protocols, and patient composition between centers may still have introduced distribution shifts. Moreover, the sample sizes at some centers were limited. The generalizability of the model therefore requires further validation in larger independent external cohorts and prospective multicenter studies. Second, the radiomic and transcriptomic data were not obtained entirely from the same patients, and some analyses therefore relied on cross-cohort integration. In addition, some imaging cohorts lacked complete RCB information, and the surrogate endpoints used may not fully correspond to actual treatment response. Future studies should establish patient-level paired cohorts with pretreatment MRI, complete RCB outcomes, and matched transcriptomic data. Third, the in silico gene perturbation analysis was based on computational inference from bulk RNA-seq data and could not fully capture cell type–specific effects, the spatial microenvironment, or complex regulatory networks. It therefore cannot replace experimental functional validation[34, 35]. Finally, future studies should combine cellular models, animal models, and paired pre- and posttreatment MRI to functionally validate **RARRES1** and its associated regulatory axes and determine whether they can drive MRI-visible tumor heterogeneity through pathway remodeling.

Availability of data and materials

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Funding

This study is supported by Shenzhen Medical Research Fund (D2501013). Guangdong Provincial Natural Science Foundation (2025A1515012849). Guangdong Provincial Medical Science and Technology Research Fund (B2025569). Guangdong Provincial and National Key Clinical Specialty Construction Project and National Key Clinical Specialty Construction Project, Sanming Project of Medicine in Shenzhen (No.SZSM202411026).

Ethics approvement

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki. The study was approved by Shenzhen Peoples Hospital ethics board (LL-KY-2021624) and individual consent for this retrospective analysis was waived. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because this research was retrospective analysis.

Author contributions

TZ: Data curation, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. GC,HZ: Data curation, Investigation. JY: Formal Analysis, Methodology, Writing – original draft. TC: Resources, Validation, Writing – review & editing. TC, JX: Formal Analysis, Methodology, Software, Supervision, Validation, Visualization, Writing – review & editing. TT,HH: Methodology, Supervision, Validation, Writing – review & editing. YY,JM: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Writing – review & editing.

1. Zaborowski AM, Wong SM. Neoadjuvant systemic therapy for breast cancer. British Journal of Surgery. 2023;110(7):765-72.
2. Tamirisa N, Hunt KK. Neoadjuvant Chemotherapy, Endocrine Therapy, and Targeted Therapy for Breast Cancer: ASCO Guideline. Ann Surg Oncol. 2022;29(3):1489-92. https://doi.org/10.1245/s10434-021-11223-3
3. Huang L, Pang D, Yang H, Li W, Wang S, Cui S, et al. Neoadjuvant–adjuvant pertuzumab in HER2-positive early breast cancer: final analysis of the randomized phase III PEONY trial. Nature Communications. 2024;15(1):2153. https://doi.org/10.1038/s41467-024-45591-7
4. Tarantino P, Hortobagyi G, Tolaney SM, Mittendorf EA. Heterogeneity of Residual Disease After Neoadjuvant Systemic Therapy in Breast Cancer: A Review. JAMA Oncol. 2024;10(11):1578-84. https://doi.org/10.1001/jamaoncol.2024.3679
5. Bottosso M, Mosele F, Michiels S, Cournède PH, Dogan S, Labaki C, et al. Moving toward precision medicine to predict drug sensitivity in patients with metastatic breast cancer. ESMO Open. 2024;9(3):102247. https://doi.org/10.1016/j.esmoop.2024.102247
6. Reig B, Lewin AA, Du L, Heacock L, Toth HK, Heller SL, et al. Breast MRI for evaluation of response to neoadjuvant therapy. Radiographics. 2021;41(3):665-79.
7. Shi Z, Huang X, Cheng Z, Xu Z, Lin H, Liu C, et al. MRI-based Quantification of Intratumoral Heterogeneity for Predicting Treatment Response to Neoadjuvant Chemotherapy in Breast Cancer. Radiology. 2023;308(1):e222830. https://doi.org/10.1148/radiol.222830
8. Su G-H, Xiao Y, You C, Zheng R-C, Zhao S, Sun S-Y, et al. Radiogenomic-based multiomic analysis reveals imaging intratumor heterogeneity phenotypes and therapeutic targets. Science advances. 2023;9(40):eadf0837.
9. Luo L, Wu M, Li M, Xin Y, Wang Q, Vardhanabhuti V, et al. A large model for non-invasive and personalized management of breast cancer from multiparametric MRI. Nature Communications. 2025;16(1):3647. https://doi.org/10.1038/s41467-025-58798-z
10. Zhang T, Zhao L, Cui T, Zhou Y, Li P, Luo C, et al. Spatial-temporal radiogenomics in predicting neoadjuvant chemotherapy efficacy for breast cancer: a comprehensive review. Journal of Translational Medicine. 2025;23(1):681.

11. You C, Su G-H, Zhang X, Xiao Y, Zheng R-C, Sun S-Y, et al. Multicenter radio-multiomic analysis for predicting breast cancer outcome and unravelling imaging-biological connection. npj Precision Oncology. 2024;8(1):193. https://doi.org/10.1038/s41698-024-00666-y
12. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. Radiology. 2021;299(2):E256. https://doi.org/10.1148/radiol.2021219005
13. Song N, Gao C, Lou X, Han Z, Wan S, He Y, et al. Deciphering the biological underpinnings behind prognostic MRI-based imaging signatures in breast cancer: a systematic review. Journal of Translational Medicine. 2025;23(1):1402. https://doi.org/10.1186/s12967-025-07341-1
14. Guo X, Xiang X, Zhuang C, Zhang H. The impact of radiogenomics on breast cancer. Am J Cancer Res. 2025;15(12):5119-39. https://doi.org/10.62347/abey9931
15. Park AY, Han M-R, Seo BK, Ju H-Y, Son GS, Lee HY, et al. MRI-based breast cancer radiogenomics using RNA profiling: association with subtypes in a single-center prospective study. Breast Cancer Research. 2023;25(1):79. https://doi.org/10.1186/s13058-023-01668-7
16. Noor H, Zheng Y, Mantz AB, Zhou R, Kozlov A, DeMartini WB, et al. A 20-feature radiomic signature of triple-negative breast cancer identifies patients at high risk of death. npj Breast Cancer. 2025;11(1):79. https://doi.org/10.1038/s41523-025-00790-3
17. Ji Y, Whitney HM, Li H, Liu P, Giger ML, Zhang X. Differences in Molecular Subtype Reference Standards Impact AI-based Breast Cancer Classification with Dynamic Contrast-enhanced MRI. Radiology. 2023;307(1):e220984. https://doi.org/10.1148/radiol.220984
18. Swain SM, Shastry M, Hamilton E. Targeting HER2-positive breast cancer: advances and future directions. Nature Reviews Drug Discovery. 2023;22(2):101-26. https://doi.org/10.1038/s41573-022-00579-0
19. Xiao J, Rahbar H, Hippe DS, Rendi MH, Parker EU, Shekar N, et al. Dynamic contrast-enhanced breast MRI features correlate with invasive breast cancer angiogenesis. npj Breast Cancer. 2021;7(1):42. https://doi.org/10.1038/s41523-021-00247-3
20. Bartsch SJ, Brožová K, Ehret V, Friske J, Fürböck C, Kenner L, et al. Non-Contrast-Enhanced Multiparametric MRI of the Hypoxic Tumor Microenvironment Allows Molecular Subtyping of Breast Cancer: A Pilot Study. Cancers (Basel). 2024;16(2). https://doi.org/10.3390/cancers16020375
21. Ramtohul T, Lepagney V, Bonneau C, Jin M, Menet E, Sauge J, et al. Use of Pretreatment Perfusion MRI–based Intratumoral Heterogeneity to Predict Pathologic Response of Triple-Negative Breast Cancer to Neoadjuvant Chemoimmunotherapy. Radiology. 2024;312(3):e240575. https://doi.org/10.1148/radiol.240575
22. Wang X, Ba R, Huang Y, Cao Y, Chen H, Xu H, et al. Time-Dependent Diffusion MRI Helps Predict Molecular Subtypes and Treatment Response to Neoadjuvant Chemotherapy in Breast Cancer. Radiology. 2024;313(1):e240288. https://doi.org/10.1148/radiol.240288

23. Kang W, Qiu X, Luo Y, Luo J, Liu Y, Xi J, et al. Application of radiomics-based multiomics combinations in the tumor microenvironment and cancer prognosis. J Transl Med. 2023;21(1):598. https://doi.org/10.1186/s12967-023-04437-4

24. Khalili N, Kazerooni AF, Familiar A, Haldar D, Kraya A, Foster J, et al. Radiomics for characterization of the glioma immune microenvironment. npj Precision Oncology. 2023;7(1):59. https://doi.org/10.1038/s41698-023-00413-9

25. Iglesia MD, Jayasinghe RG, Chen S, Terekhanova NV, Herndon JM, Storrs E, et al. Differential chromatin accessibility and transcriptional dynamics define breast cancer subtypes and their lineages. Nature Cancer. 2024;5(11):1713-36. https://doi.org/10.1038/s43018-024-00773-6

26. Orlando KA, Grimm SA, Wade PA. Pioneering new enhancers by GATA3: role of facilitating transcription factors and chromatin remodeling. Nucleic Acids Res. 2025;53(11). https://doi.org/10.1093/nar/gkaf473

27. Turova P, Kushnarev V, Baranov O, Butusova A, Menshikova S, Yong ST, et al. The Breast Cancer Classifier refines molecular breast cancer classification to delineate the HER2-low subtype. NPJ Breast Cancer. 2025;11(1):19. https://doi.org/10.1038/s41523-025-00723-0

28. Martin EM, Orlando KA, Yokobori K, Wade PA. The estrogen receptor/GATA3/FOXA1 transcriptional network: lessons learned from breast cancer. Curr Opin Struct Biol. 2021;71:65-70. https://doi.org/10.1016/j.sbi.2021.05.015

29. Aquino-Acevedo AN, Orengo-Orengo JA, Cruz-Robles ME, Saavedra HI. Mitotic kinases are emerging therapeutic targets against metastatic breast cancer. Cell Division. 2024;19(1):21. https://doi.org/10.1186/s13008-024-00125-x

30. Groelly FJ, Fawkes M, Dagg RA, Blackford AN, Tarsounas M. Targeting DNA damage response pathways in cancer. Nat Rev Cancer. 2023;23(2):78-94. https://doi.org/10.1038/s41568-022-00535-5

31. Sinha D, Nag P, Nanayakkara D, Duijf PHG, Burgess A, Raninga P, et al. Cep55 overexpression promotes genomic instability and tumorigenesis in mice. Communications Biology. 2020;3(1):593. https://doi.org/10.1038/s42003-020-01304-6

32. Liu Z, Duan T, Zhang Y, Weng S, Xu H, Ren Y, et al. Radiogenomics: a key component of precision cancer medicine. British Journal of Cancer. 2023;129(5):741-53. https://doi.org/10.1038/s41416-023-02317-8

33. Huang G, Du S, Gao S, Guo L, Zhao R, Bian X, et al. Molecular subtypes of breast cancer identified by dynamically enhanced MRI radiomics: the delayed phase cannot be ignored. Insights Imaging. 2024;15(1):127. https://doi.org/10.1186/s13244-024-01713-9

34. Ahlmann-Eltze C, Huber W, Anders S. Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nature Methods. 2025;22(8):1657-61. https://doi.org/10.1038/s41592-025-02772-6

35. Zhu O, Li J. Scouter predicts transcriptional responses to genetic perturbations with large language model embeddings. Nature Computational Science. 2026;6(1):21-8. https://doi.org/10.1038/s43588-025-00912-8

Table 1 Patient characteristic.

	Overall				Center1				Center 2			
	Training set		Validation set		Training set		Validation set		Training set		Validation set	
	RCB 0/1	RCB 2/3	RCB 0/1	RCB 2/3	RCB 0/1	RCB 2/3	RCB 0/1	RCB 2/3	RCB 0/1	RCB 2/3	RCB 0/1	RCB 2/3
Age												

Age≤50	76 (19.34%)	66 (16.79%)	20 (20.20%)	15 (15.15%)	52 (16.46%)	62 (19.62%)	15 (18.99%)	15 (18.99%)	24 (31.17%)	4 (5.19%)	5 (25.00%)	0 (0.00%)
Age>50	162 (41.22%)	89 (22.65%)	41 (41.41%)	23 (23.23%)	124 (39.24%)	78 (24.68%)	27 (34.18%)	22 (27.85%)	38 (49.35%)	11 (14.29%)	14 (70.00%)	1 (5.00%)
T												
T1	28 (7.12%)	11 (2.80%)	8 (8.08%)	0 (0.00%)	19 (6.01%)	8 (2.53%)	5 (6.33%)	0 (0.00%)	9 (11.69%)	3 (3.90%)	3 (15.00%)	0 (0.00%)
T2	178 (45.29%)	106 (26.97%)	42 (42.42%)	31 (31.31%)	138 (43.67%)	98 (31.01%)	27 (34.18%)	30 (37.97%)	40 (51.95%)	8 (10.39%)	15 (75.00%)	1 (5.00%)
T3	23 (5.85%)	21 (5.34%)	8 (8.08%)	6 (6.06%)	17 (5.38%)	19 (6.01%)	8 (10.13%)	6 (7.59%)	6 (7.79%)	2 (2.60%)	0 (0.00%)	0 (0.00%)
T4	9 (2.29%)	17 (4.33%)	3 (3.03%)	1 (1.01%)	2 (0.63%)	15 (4.75%)	2 (2.53%)	1 (1.27%)	7 (9.09%)	2 (2.60%)	1 (5.00%)	0 (0.00%)
N												
N0	99 (25.19%)	45 (11.45%)	22 (22.22%)	14 (14.14%)	79 (25.00%)	42 (13.29%)	16 (20.25%)	14 (17.72%)	20 (25.97%)	3 (3.90%)	6 (30.00%)	0 (0.00%)
N1	113 (28.75%)	70 (17.81%)	36 (36.36%)	22 (22.22%)	81 (25.63%)	61 (19.30%)	24 (30.38%)	21 (26.58%)	32 (41.56%)	9 (11.69%)	12 (60.00%)	1 (5.00%)
N2	13 (3.31%)	24 (6.11%)	2 (2.02%)	0 (0.00%)	6 (1.90%)	23 (7.28%)	1 (1.27%)	0 (0.00%)	7 (9.09%)	1 (1.30%)	0 (0.00%)	0 (0.00%)
N3	10 (2.54%)	13 (3.31%)	0 (0.00%)	2 (2.02%)	7 (2.22%)	11 (3.48%)	0 (0.00%)	2 (2.53%)	3 (3.90%)	2 (2.60%)	1 (5.00%)	0 (0.00%)
M												
M0	235 (59.80%)	153 (38.93%)	60 (60.61%)	37 (37.37%)	174 (55.06%)	138 (43.67%)	41 (51.90%)	36 (45.57%)	61 (79.22%)	15 (19.48%)	19 (95.00%)	1 (5.00%)
M1	1 (0.25%)	1 (0.25%)	1 (1.01%)	0 (0.00%)	0 (0.00%)	1 (0.32%)	1 (1.27%)	0 (0.00%)	1 (1.30%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
ER												
Negative	132 (33.59%)	59 (15.01%)	23 (23.23%)	11 (11.11%)	108 (34.18%)	56 (17.72%)	21 (26.58%)	11 (13.92%)	24 (31.17%)	3 (3.90%)	2 (10.00%)	0 (0.00%)
Positive	106 (26.97%)	96 (24.43%)	38 (38.38%)	27 (27.27%)	115 (36.39%)	67 (21.20%)	21 (26.58%)	26 (32.91%)	38 (49.35%)	12 (15.58%)	17 (85.00%)	1 (5.00%)
PR												
Negative	139 (35.37%)	71 (18.07%)	34 (34.34%)	13 (13.13%)	115 (36.39%)	67 (21.20%)	29 (36.71%)	12 (15.19%)	24 (31.17%)	4 (5.19%)	5 (25.00%)	1 (5.00%)
Positive	99 (25.19%)	84 (21.37%)	27 (27.27%)	25 (25.25%)	61 (19.30%)	73 (23.10%)	13 (16.46%)	25 (31.65%)	38 (49.35%)	11 (14.29%)	14 (70.00%)	0 (0.00%)
HER2												

Negative	91 (23.16%)	101 (25.70%)	21 (21.21%)	26 (26.26%)	64 (20.25%)	94 (29.75%)	9 (11.39%)	25 (31.65%)	27 (35.06%)	7 (9.09%)	12 (60.00%)	1 (5.00%)
Positive	147 (37.40%)	54 (13.74%)	40 (40.40%)	12 (12.12%)	112 (35.44%)	46 (14.56%)	33 (41.77%)	12 (15.19%)	35 (45.45%)	8 (10.39%)	7 (35.00%)	0 (0.00%)

Table 2. Clinical characteristics of matched I-SPY

Characteristic	**Level**	**Overall**	**RCB-0/I**	**RCB-II/III**
Age, years	Median (IQR); mean ± SD	44.20 (39.40-51.40); 46.06 ± 8.83	40.80 (38.07-42.65); 41.31 ± 6.68	45.90 (40.65-52.85); 47.47 ± 8.96
DRFS follow-up, years	Median (IQR); mean ± SD	2.48 (1.64-3.26); 2.54 ± 1.26	2.53 (2.37-3.59); 2.80 ± 1.28	2.45 (1.61-3.25); 2.46 ± 1.26
Clinical AJCC stage	IIA	13 (21.3)	3 (21.4)	10 (21.3)
	IIB	18 (29.5)	6 (42.9)	12 (25.5)
	IIIA	22 (36.1)	5 (35.7)	17 (36.2)
	IIIB	4 (6.6)	0 (0.0)	4 (8.5)
	IIIC	2 (3.3)	0 (0.0)	2 (4.3)
	Inflammatory	2 (3.3)	0 (0.0)	2 (4.3)
Clinical T stage	T1	1 (1.6)	0 (0.0)	1 (2.1)
	T2	27 (44.3)	8 (57.1)	19 (40.4)
	T3	28 (45.9)	6 (42.9)	22 (46.8)
	T4	5 (8.2)	0 (0.0)	5 (10.6)
Clinical nodal status	N0	18 (29.5)	5 (35.7)	13 (27.7)
	N1	36 (59.0)	8 (57.1)	28 (59.6)
	N2	5 (8.2)	1 (7.1)	4 (8.5)
	N3	2 (3.3)	0 (0.0)	2 (4.3)
Histologic grade	Missing	5 (8.2)	1 (7.1)	4 (8.5)
	1	4 (6.6)	0 (0.0)	4 (8.5)
	2	18 (29.5)	3 (21.4)	15 (31.9)
	3	23 (37.7)	7 (50.0)	16 (34.0)
	4=Indeterminate	11 (18.0)	3 (21.4)	8 (17.0)
ER status by IHC	Negative	28 (45.9)	11 (78.6)	17 (36.2)
	Positive	31 (50.8)	3 (21.4)	28 (59.6)
	Indeterminate	1 (1.6)	0 (0.0)	1 (2.1)
	Missing	1 (1.6)	0 (0.0)	1 (2.1)

PR status by IHC	Negative	30 (49.2)	11 (78.6)	19 (40.4)
	Positive	28 (45.9)	3 (21.4)	25 (53.2)
	Indeterminate	2 (3.3)	0 (0.0)	2 (4.3)
	Missing	1 (1.6)	0 (0.0)	1 (2.1)
HER2 status	Negative	45 (73.8)	10 (71.4)	35 (74.5)
	Positive	3 (4.9)	1 (7.1)	2 (4.3)
	Indeterminate	4 (6.6)	0 (0.0)	4 (8.5)
	Missing	9 (14.8)	3 (21.4)	6 (12.8)
PAM50 subtype	Basal	30 (49.2)	10 (71.4)	20 (42.6)
	Her2	4 (6.6)	1 (7.1)	3 (6.4)
	LumA	14 (23.0)	1 (7.1)	13 (27.7)
	LumB	7 (11.5)	1 (7.1)	6 (12.8)
	Normal	6 (9.8)	1 (7.1)	5 (10.6)
Pathologic response	pCR	13 (21.3)	13 (92.9)	0 (0.0)
	RD	48 (78.7)	1 (7.1)	47 (100.0)
RCB class	RCB-0/I	14 (23.0)	14 (100.0)	0 (0.0)
	RCB-II	31 (50.8)	0 (0.0)	31 (66.0)
	RCB-III	16 (26.2)	0 (0.0)	16 (34.0)
RCB 0/I prediction label	RCB-0/I	18 (29.5)	14 (100.0)	4 (8.5)
	RCB-II/III	43 (70.5)	0 (0.0)	43 (91.5)
DRFS event	Censored	49 (80.3)	14 (100.0)	35 (74.5)
	DRFS event	12 (19.7)	0 (0.0)	12 (25.5)

Figure 1. Receiver Operating Characteristic Curves of the RCB Prediction Model in the Training and Test Sets Across Two Centers

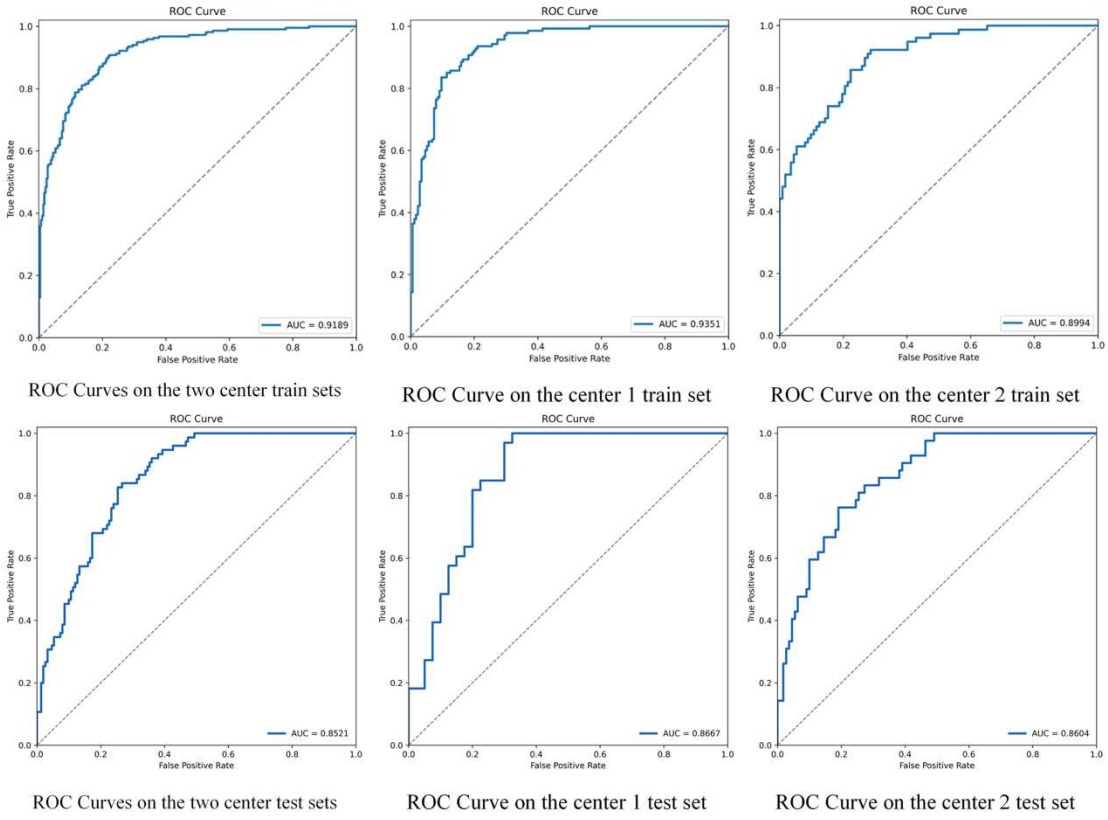
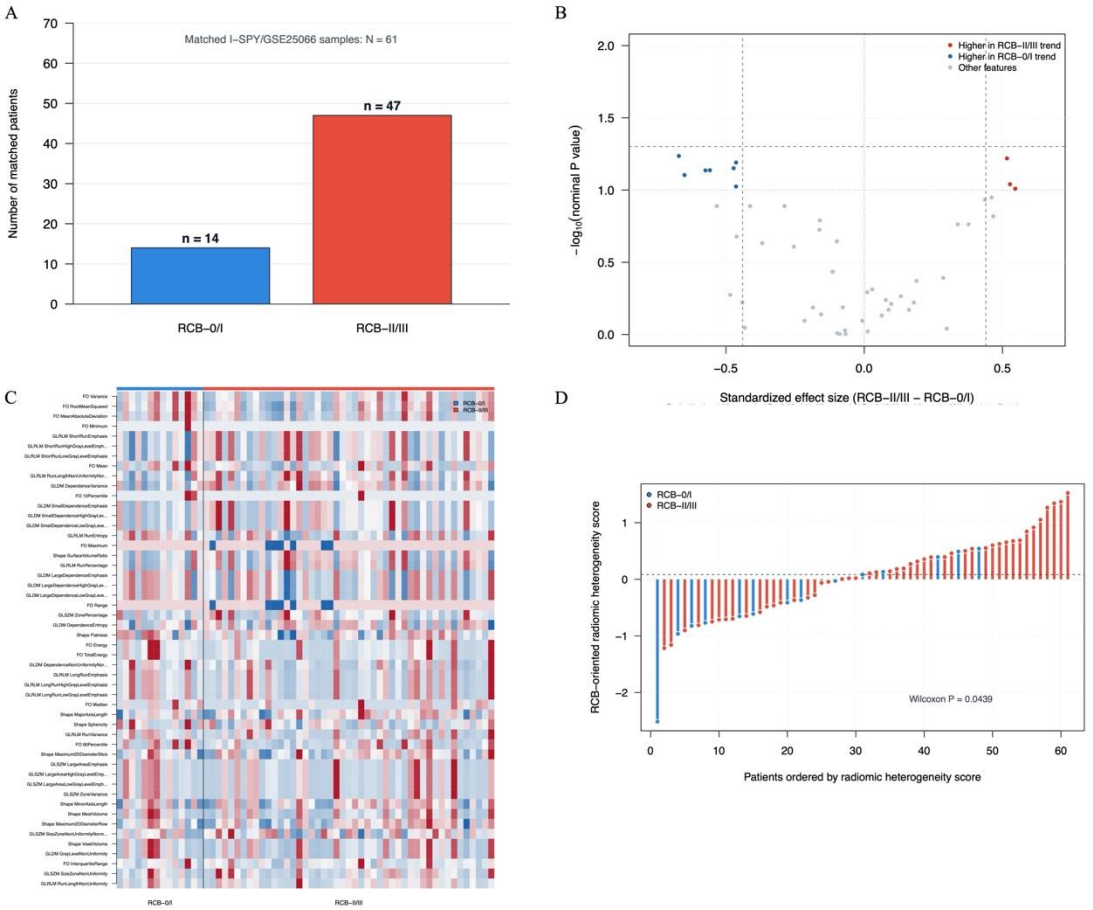


Figure 2. Pretreatment MRI radiomic phenotypes associated with residual cancer burden.



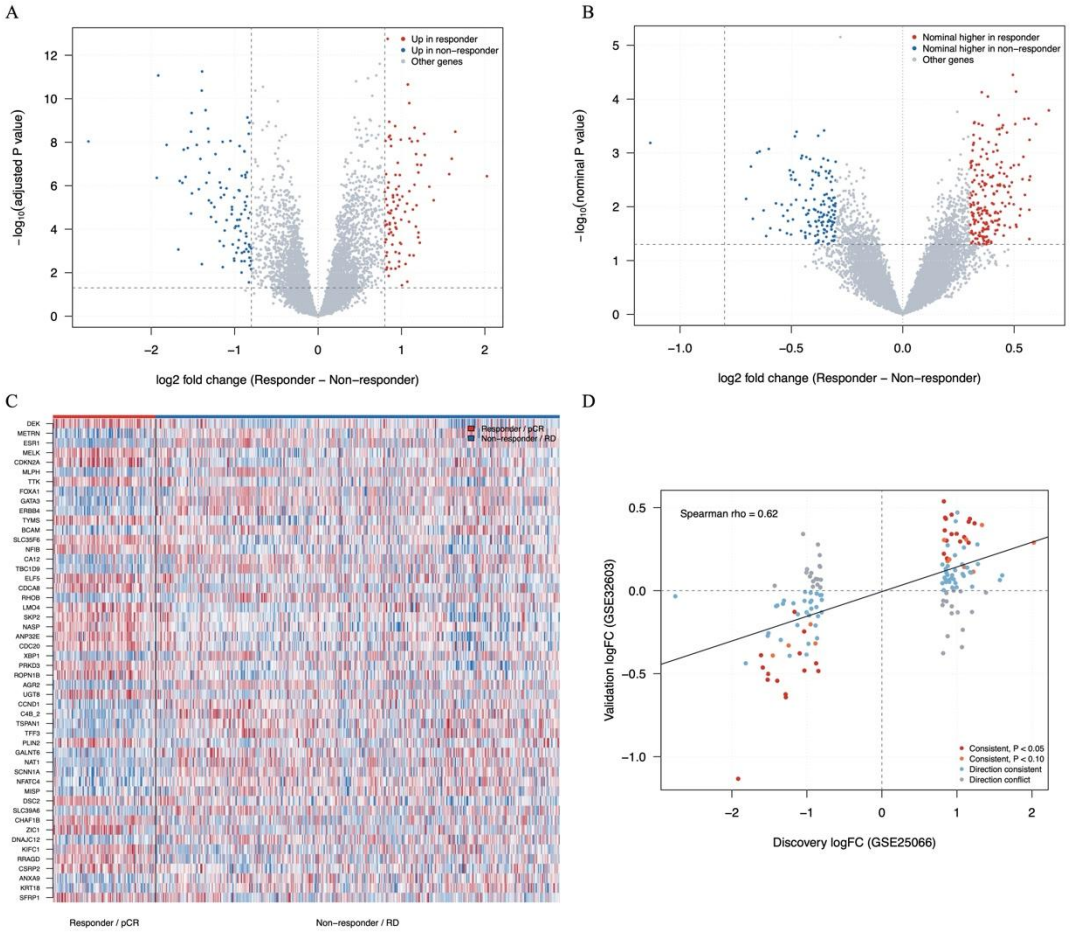
(A) Distribution of RCB groups in the matched I-SPY/GSE25066 cohort. RCB-0/I and RCB-II/III represent low and high residual burden, respectively.

(B) Volcano plot of RCB-associated radiomic features. The x-axis shows the standardized effect size (RCB-II/III minus RCB-0/I), and the y-axis shows −log10 nominal P value. Red and blue indicate features enriched in RCB-II/III and RCB-0/I, respectively; gray indicates other features. No feature remained significant after FDR correction.

(C) Heatmap of the top-ranked RCB-associated radiomic features. Patients are ordered by radiomic score and grouped by RCB status.

(D) Distribution of the RCB-oriented radiomic heterogeneity score across matched patients. The score was nominally associated with RCB group.

Figure 3. Discovery and validation of treatment response–associated differentially expressed genes.



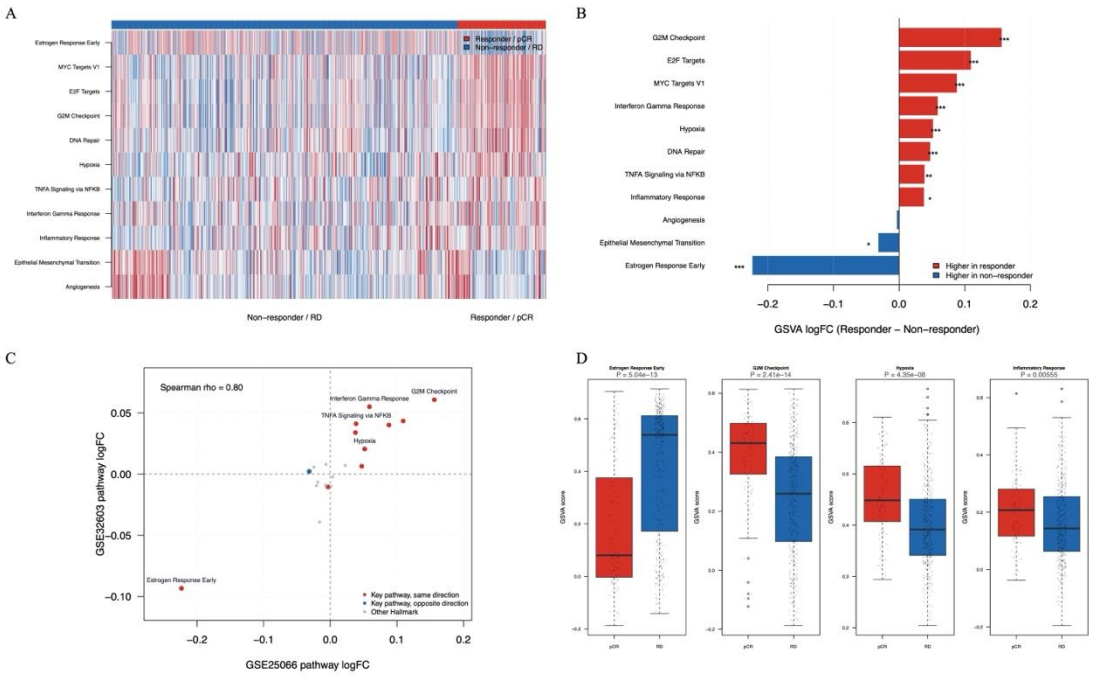
(A) Volcano plot of differential expression in the GSE25066 discovery cohort. The x-axis shows log2 fold change (responder minus nonresponder), and the y-axis shows −log10 adjusted *P* value. Red and blue indicate genes upregulated in responders and nonresponders, respectively; gray indicates other genes.

(B) Validation volcano plot for GSE32603 showing nominal *P*-value trends, as no gene reached FDR significance.

(C) Heatmap of the top 50 discovery DEGs, showing expression patterns across responder/pCR and nonresponder/RD samples.

(D) Comparison of logFC values between GSE25066 and GSE32603 for detectable discovery DEGs. Colors indicate directional concordance and nominal validation support.

Figure 4. Pathway-level programs associated with neoadjuvant treatment response.



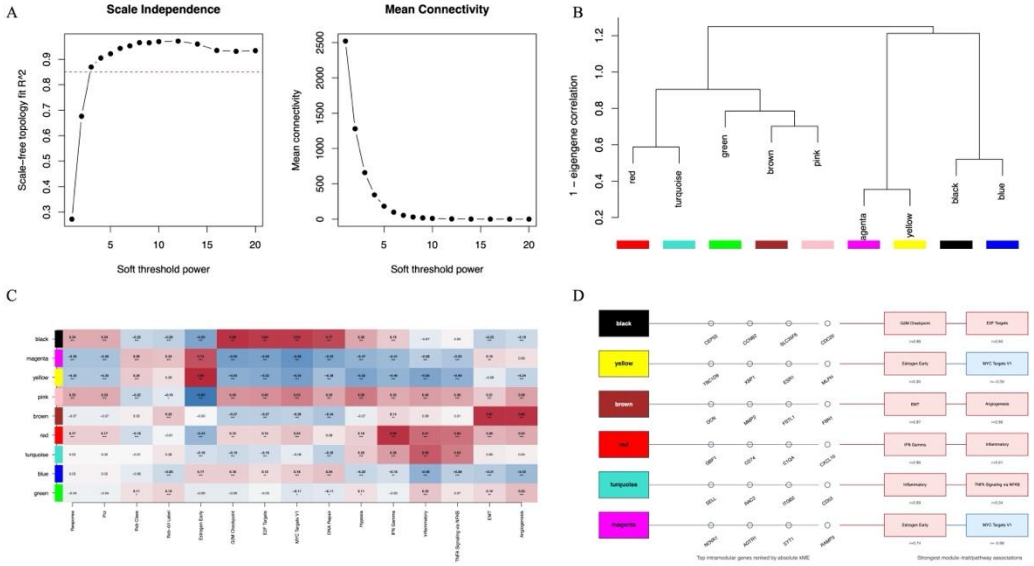
(A) Heatmap of GSVA/ssGSEA scores for selected MSigDB Hallmark pathways in the GSE25066 discovery cohort. Patients are grouped by response status. Red and blue indicate higher and lower row-standardized pathway activity, respectively.

(B) Differential pathway activity in GSE25066. LogFC was calculated as responder/pCR minus nonresponder/RD. Red and blue indicate pathways enriched in responders and nonresponders, respectively. Asterisks denote significance after adjustment for multiple testing.

(C) Comparison of pathway logFC values between GSE25066 and GSE32603. Red and blue indicate concordant and discordant directions among key pathways, respectively; gray indicates other Hallmark pathways.

(D) Activity distributions of representative pathways, including Estrogen Response Early, G2M Checkpoint, Hypoxia, and Inflammatory Response. P values were calculated using Wilcoxon rank-sum tests.

Figure 5. WGCNA identifies co-expression modules associated with treatment response, RCB, and pathway activity.



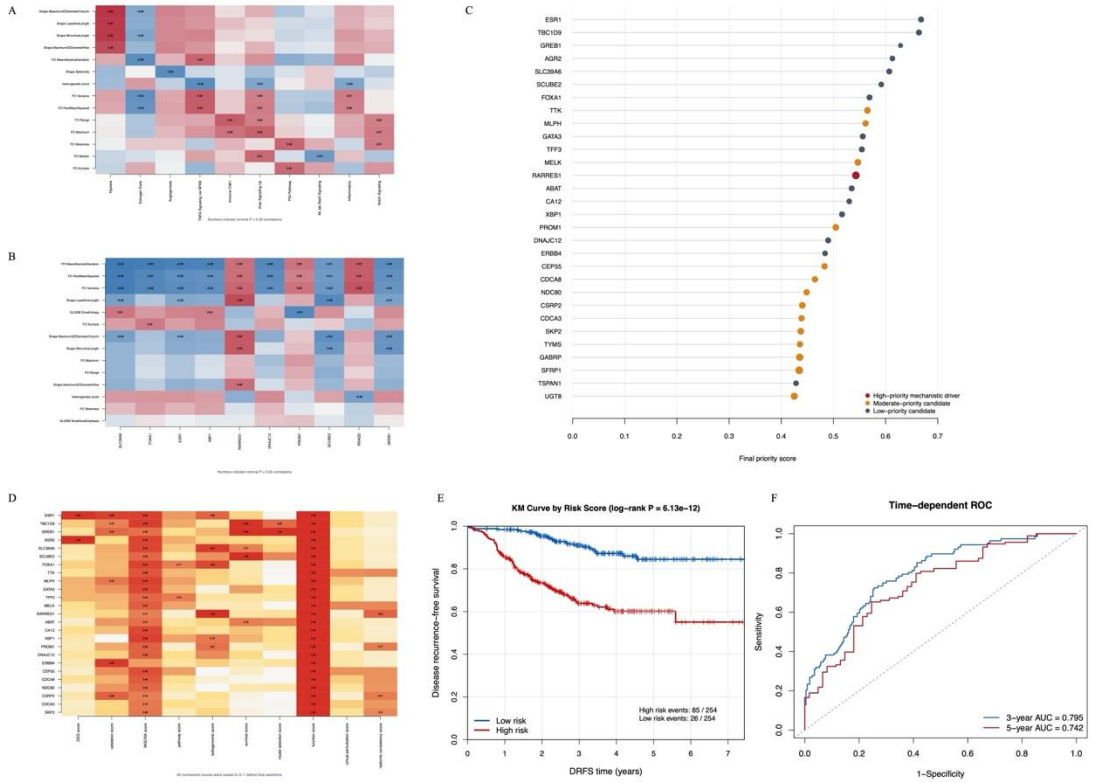
(A) Selection of the soft-thresholding power for the signed WGCNA network in GSE25066. Based on scale-free topology fit and mean connectivity, a power of 3 was selected.

(B) Module eigengene dendrogram based on eigengene correlations. Colors indicate co-expression modules.

(C) Correlations of module eigengenes with treatment response, pCR, RCB-related variables, and selected Hallmark pathway scores. Cells show Pearson correlation coefficients; red and blue indicate positive and negative correlations, respectively. Asterisks denote significance after multiple-testing correction.

(D) Simplified network of key modules, representative hub genes, and major trait/pathway associations. Hub genes were ranked by absolute module membership (kME); red and blue edges indicate positive and negative correlations, respectively.

Figure 6. Integrated radiogenomic analysis, hub-gene prioritization, and prognostic evaluation.



(A) Spearman correlations between selected MRI radiomic features and GSVA pathway scores. Red and blue indicate positive and negative correlations, respectively; values denote nominally significant correlations ($P < 0.05$). Shape features were positively associated with Hypoxia, whereas the radiomic heterogeneity score was negatively associated with TNF/NF-κB Signaling.

(B) Correlations between radiomic features and candidate hub-gene expression. Intensity-dispersion features were associated with luminal/endocrine genes, whereas shape features were associated with genes including RARRES1 and SCUBE2.

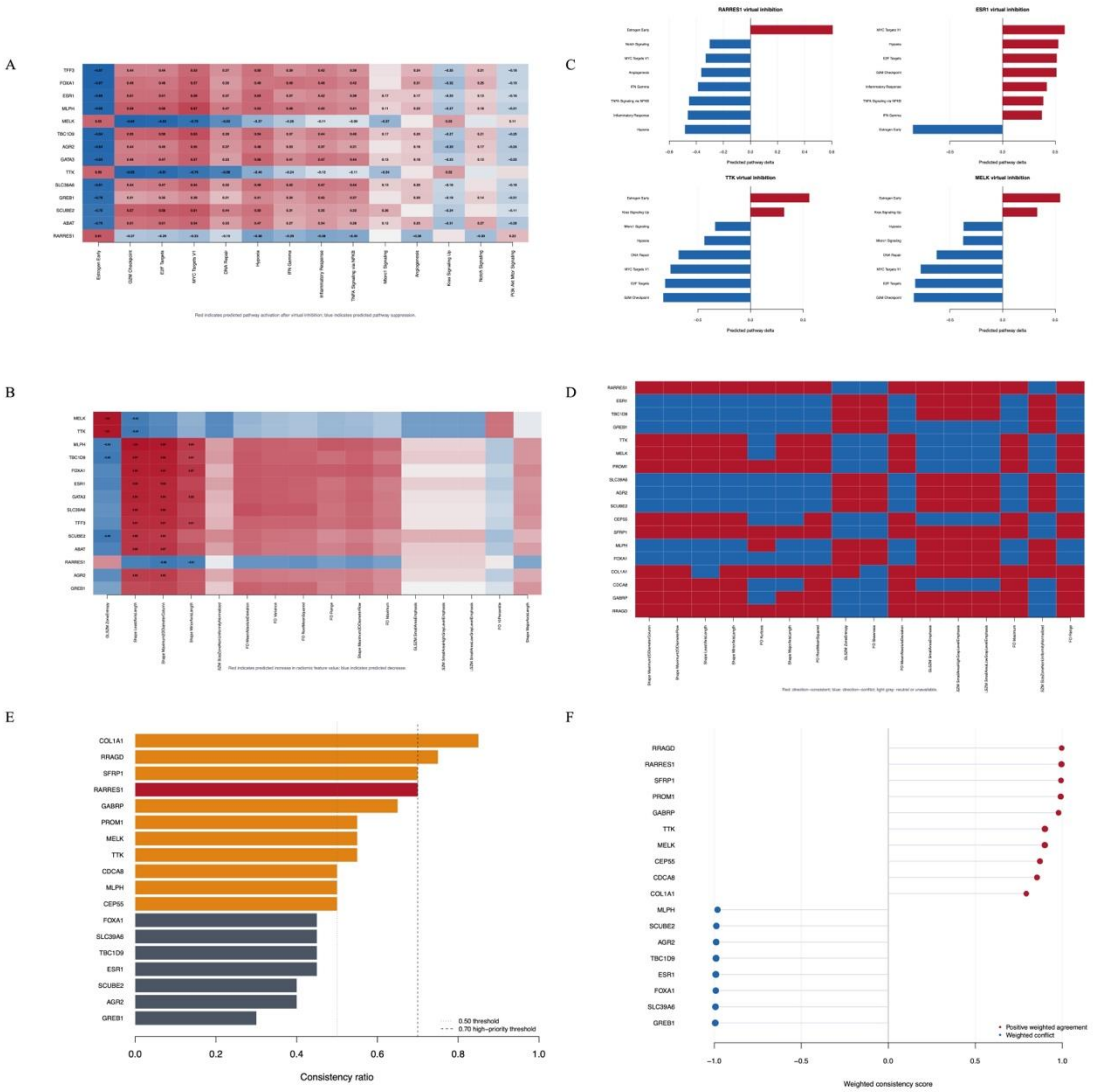
(C) Final priority scores of the top candidate genes. Color indicates priority level, and dot size represents the radiomic Consistency Ratio.

(D) Normalized evidence scores for the top candidate genes, integrating differential expression, validation, WGCNA, pathway and radiogenomic associations, survival analysis, model selection, functional annotation, in silico perturbation, and radiomic consistency.

(E) Kaplan–Meier curves comparing distant recurrence-free survival (DRFS) between multigene signature–defined risk groups.

(F) Time-dependent ROC curves evaluating the prognostic risk score for predicting DRFS.

Figure 7. In silico gene perturbation links pathway remodeling to predicted MRI phenotype shifts and radiomic concordance.



(A) Predicted pathway changes following inhibition of top-ranked candidate genes. Red and blue indicate pathway activation and suppression, respectively; values denote predictions supported by nominal gene–pathway correlations ($P < 0.05$).

(B) Representative perturbation profiles for RARRES1, ESR1, TTK, and MELK, showing the most strongly altered pathways. Positive and negative values indicate predicted activation and suppression, respectively.

(C) Predicted gene-level changes in MRI radiomic features. Each gene–feature shift was calculated by summing the products of predicted pathway changes and corresponding pathway–radiomic correlations. Red and blue indicate predicted increases and decreases, respectively.

(D) Concordance between predicted radiomic changes and observed differences between high- and low-heterogeneity groups. Red indicates concordance, blue indicates directional conflict, and gray indicates neutral or unavailable comparisons.

(E) Radiomic Consistency Ratios for representative genes. The dashed line indicates the prespecified threshold of 0.70 for high-priority mechanistic drivers.

(F) Weighted Consistency Scores for representative genes. Positive and negative values indicate weighted concordance and directional conflict, respectively. All perturbation results are computational approximations based on bulk RNA-seq and should not be interpreted as experimental gene-knockout effects.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.docx