Radiogenomics Analysis Reveals Tumor Heterogeneity of Triple-Negative Breast Cancer

Lin Jiang
  Fudan University Shanghai Cancer Center

Chao You
  Fudan University Shanghai Cancer Center

Yi Xiao
  Fudan University Shanghai Cancer Center

He Wang
  Fudan University

Bing Qing Xia
  Shanghai Jiao Tong University

Ren Cheng Zheng
  Fudan University

Dan Dan Zhang
  Fudan University

Yi-Zhou Jiang
  Fudan University Shanghai Cancer Center

Yajia Gu
  Fudan University Shanghai Cancer Center

Zhi-Ming Shao (✉ zhimin_shao@yeah.net)
  Fudan University Shanghai Cancer Center

Article

Abstract

Triple-negative breast cancer (TNBC) is a subset of breast cancer with adverse prognosis and significant tumor heterogeneity. Here, we used MRI images from a breast cancer cohort consisting of 860 patients to construct a radiomic signature that could identify TNBC with an AUC of 0.92 (95% CI: 0.887 to 0.953) and validated in another cohort. Moreover, we developed radiomic signatures to distinguish TNBC subtypes with moderate efficacy. Furthermore, we identified peritumoral dependence nonuniformity of the gray level dependence matrix, which captures the intratumor heterogeneity in the tumor boundary, as the most significant prognostic factor (P = 0.04 for recurrence-free survival and P = 0.02 for overall survival). The integration of transcriptomic and metabolomic data indicated that high peritumor heterogeneity was related to immune suppression and enhanced metabolism. Our findings suggest that radiomics data can serve as a noninvasive predictor for molecular subtyping and clinical outcome in patients with TNBC.

Introduction

Breast cancers that lack expression of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) are classified as triple-negative breast cancers (TNBCs) and comprise 15% to 20% of newly diagnosed breast cancers (1, 2). TNBCs are characterized by aggressive biological behavior, a high incidence of relapse, and unfavorable prognosis (3, 4). Recent years have witnessed increasing recognition that TNBC is a highly heterogeneous disease (5-7) and it thus requires the identification of subtype-specific therapeutic targets (8-11). With the largest multiomics database to date, our previous work unveiled the genomic and transcriptomic landscape of 465 Chinese TNBC patients and classified TNBCs into four transcriptomic subtypes with distinct characteristics: 1) basal-like immune-suppressed (BLIS), 2) immunomodulatory (IM), 3) mesenchymal-like (MES), and 4) luminal androgen receptor (LAR).

In the past decade, radiomics has been an emerging field that transforms medical images into mineable data by acquiring multiple quantitative image features (12, 13). Compared to conventional invasive biopsies, the radiomic approach has several advantages. First, radiomics is a non-invasive method to infer tumor characteristics and can be carried out several times during treatment (11, 14, 15). Moreover, compared to genomic sequencing, which selects only a small part of the tumors, radiomics elucidates the landscape of a tumor and is not subjected to selection bias, and thus could comprehensively explore tumor heterogeneity (16-18). Previous studies focused on radiomic texture analysis have quantified tumor heterogeneity and suggested its associations with unfavorable prognosis in breast cancer (19, 20). These results warrant further research into evaluating tumor heterogeneity via a radiomic approach.

However, a multiomics TNBC dataset containing radiomics data with a large sample size has yet to be reported, and the correlation between radiomic features and genomic alterations remains largely unknown. In this study, we performed radiomic profiling based on contrast-enhanced magnetic resonance imaging (CE-MRI) images from 860 Chinese breast cancer patients, which composed the largest breast cancer MRI radiomics dataset to date. This cohort comprised 246 TNBC patients, including 202 samples

with genomic, transcriptomic or metabolomic data, and aimed to build a radiomics model for non-invasive TNBC classification and patient outcome stratification. We also integrated radiomic data with our transcriptomic, metabolomic and clinical data to illustrate the biological basis of prognostic radiomic features in TNBC.

Results

Identification of TNBC from the Whole Breast Cancer Cohort via a Radiomics Approach

To explore whether a non-invasive radiomics approach could be used to identify TNBCs from non TNBCs, we first established a FUSCC breast cancer radiomics cohort that included all subtypes of breast cancer. A total of 860 patients from FUSCC with baseline breast CE-MRI images for radiomics analysis between August 2009 and June 2019 were retrospectively enrolled (**Figure 1A**). Hormone receptor positivity was observed in 468 patients, HER-2 overexpression was observed in 268 patients, and 246 patients had TNBC (**Figure 1B**). This radiomic cohort was further randomly divided into 50% training and 50% validation cohorts (**Figure 1A**). Another 54 patients from IPMCH were recruited as an independent external validation cohort. Four types of ROIs were employed: tumoral, peritumoral, intratumoral and tumor-peritumoral. To the best of our knowledge, our cohort provides the largest breast cancer radiomic dataset to date (**Figure 1C**). The analysis plan of this radiogenomics study is shown in **Figure 1D**.

Using the training set in the FUSCC breast cancer radiomics cohort, we developed radiomic prediction models through feature selection and model building. Hormone and HER-2 receptor status determined by IHC and fluorescence in situ hybridization were regarded as the ground truth to evaluate the prediction efficacy. Using the 10-fold cross validation LASSO model ($\alpha = 0$), 11 variables were retained for the development of TNBC prediction in the training cohort (**Supplementary Figure 1**). These radiomic features are presented in **Supplementary Table 1**. Logistic regression (LR) and support vector machine (SVM) were applied to establish prediction models based on retained features, and radiomics scores were calculated for each patient in both the training and FUSCC validation cohorts using these two methods.

Using the LR-based model, the ability of the radiomic signature to classify TNBC versus non-TNBC was shown to have an AUC of 0.969 (95% CI: 0.951-0.987) in the training cohort and an AUC of 0.92 (95% CI: 0.887-0.953) in the FUSCC validation cohort, while the SVM-based model demonstrated an AUC of 0.968 (95% CI: 0.95−0.986) in the training cohort and an AUC of 0.922 (95% CI: 0.89-0.954) in the FUSCC validation cohort (**Figure 2A**). We further calculated that the Jaccard similarity index was 0.881 and 0.894 in the validation cohort for the LR and SVM models, respectively. To validate the effectiveness of this radiomic signature in images from different MRI machines, we calculated the accuracy of these prediction models separately according to MRI machines. To expand the usage of this radiomics signature to a larger extent, we further validated the efficacy of the prediction models in the IPMCH validation cohort. The LR and SVM models yielded AUCs of 0.723 (95% CI: 0.552-0.894) and 0.704 (95% CI: 0.534-0.874), respectively (**Figure 2B**). Among these retained features, the mean of tumor zone entropy, which characterizes the mean heterogeneity during different MRI sequences, was significantly

higher in TNBC (**Figure 2C**, P < 0.001), and post hoc analysis indicated that tumor heterogeneity was highly distinctive in identifying TNBC patients, while tumor heterogeneity was not significantly different between luminal and HER2-enriched tumors. In summary, these results implied a satisfactory separation of TNBC versus non-TNBC via a radiomics approach and suggested that high intratumoral heterogeneity was a crucial feature of TNBC.

Predictive Value of Radiomics in Distinguishing TNBC Molecular Subtypes

Since the radiomic signature could identify TNBC with high accuracy, we further explored whether radiomic signatures could distinguish different TNBC molecular subtypes. We employed CE-MRI images from 202 patients who were enrolled in the FUSCC breast cancer radiomics cohort as well as the FUSCC TNBC cohort, which is the largest TNBC multiomics cohort to date (n = 465), to validate this hypothesis. The baseline characteristics of this subcohort are shown in **Supplementary Table 2**. A total of 167 samples with radiomic data also had transcriptomic data, and 138 samples also had metabolomic data (**Supplementary Figure 2**). Transcriptomic TNBC subtypes were regarded as the ground truth to evaluate the prediction efficacy(7).

LASSO and Student's t test retained 4, 11, 2 and 7 radiomic features that were most relevant to BLIS, IM, MES and LAR subtypes (**Supplementary Figure 3A-3D**) in the training cohort. These features are presented in **Supplementary Table 3**. LR and SVM were used to construct prediction models in the training cohort and validation cohort based on the selected features. Radiomics scores were calculated for each patient using these three methods.

The AUCs and confidence intervals of the prediction models for each TNBC subtype using different methods are shown in **Figure 3A-3D** and **Supplementary Table 4**. In the FUSCC validation cohort, distinguishing the MES subtype (based on the LR model) yielded an AUC of 0.796 (95% CI: 0.65-0.941), and the prediction models of the BLIS (based on the SVM model) and IM (based on the LR model) subtypes obtained AUCs of 0.719 (95% CI: 0.57-0.867) and 0.669 (95% CI: 0.481-0.858), respectively, while the AUC of identifying the LAR (based on the SVM model) subtype only reached 0.598 (95% CI: 0.416-0.781). As a retained feature of the MES prediction model, tumor elongation was significantly larger in the MES subtype than in the other TNBC molecular subtypes (*P* = 0.021 in the training cohort and *P* = 0.019 in the FUSCC validation cohort, **Supplementary Figure 4A-4B**). This easy-to-use radiomic feature could facilitate the classification of MES subtype in clinical practice.

A previous study developed an IHC-based method that specifically used the staining results of AR, CD8A, FOXC1 and DCLK1 to classify TNBC molecular subtypes (21). We further explored the discriminatory power of prediction models combining radiomic features and IHC data. Since IHC alone could identify the LAR subtype with outstanding efficacy (AUC = 0.932) (21) and a satisfactory radiomic model was established for MES subtype identification, we built combined signatures for identifying BLIS and IM subtypes. For predicting BLIS subtype, the SVM-based combined prediction model reached an AUC of 0.975 (95% CI, 0.906-1) in the validation cohort (**Figure 3E-3F**). For predicting IM subtype, the LR-based combined prediction model displayed an AUC of 0.731 (95% CI, 0.373-1) in the

validation cohort (**Figure 3G-3H**). Combined signatures showed better performance in the BLIS subtype than individual IHC and radiomics-based models, but no statistical significance was found in the IM subtype. The limited sample size (n = 42 in the training cohort and n = 14 in the validation cohort with radiomics, transcriptomics and IHC staining data) restricted the performance of this combined signature and warrants further research with a larger sample size.

Prognostic Value of Peritumoral Heterogeneity Derived from Radiomics

With the detailed clinical follow-up data of our TNBC cohort, we evaluated robust prognostic radiomic features. According to stringent filtering criteria, variance among the MRI sequences of dependence nonuniformity extracted from peritumoral ROIs (Peri_V_DN), a feature from the gray level dependence matrix group, was identified (**Figure 4A**). Typical breast CE-MRI images with high and low Peri_V_DN values are shown in **Figure 4B**. The stratification of patients with survival differences using the median value as the cutoff was verified in both the training cohort and validation cohort (**Figure 4C**). The multivariate Cox proportional hazards model also revealed that low Peri_V_DN independently predicted better recurrence-free survival (RFS) and overall survival (OS) in TNBC patients (**Table 1**). As a risk-increasing feature, Peri_V_DN represents the variation pattern of peritumoral heterogeneity through different sequences, with a lower value indicating less change in peritumoral heterogeneity among the sequences of the image.

Moreover, we systematically analyzed the correlation between the Peri_V_DN value and tumor characteristics. We observed larger tumor sizes and more pathologically confirmed metastatic lymph nodes ($P = 1.8 \times 10^{-8}$ and $P = 0.00661$, respectively) in the high Peri_V_DN group than in the low Peri_V_DN group (**Figure 4D**). The high Peri_V_DN group included more patients with the BLIS subtype, and the low Peri_V_DN group comprised more patients with the IM subtype ($P = 0.02$), while the distribution of the PAM50 subtypes was balanced (**Figure 4E**). We analyzed the correlation between the Peri_V_DN value and TNBC microenvironment clusters according to TNBC microenvironment subtypes (22). The results revealed a tendency for the high Peri_V_DN group to consist of more "immune-desert" cluster 1 tumors, while the low Peri_V_DN group included more "immune-inflamed" cluster 3 tumors ($P = 0.09$, **Figure 4E**). Fibrosis and necrosis grade evaluated by H.E. sections showed no difference between the two Peri_V_DN groups (**Supplementary Figure 5A**). Other molecular biomarkers that putatively predict TNBC precision treatment efficacy, including stromal tumor-infiltrating lymphocytes (TILs), IHC CD8 readings, tumor mutation burden (TMB) and homologous recombination deficiency (HRD) score, displayed balanced distributions between the Peri_V_DN groups as well (**Supplementary Figure 5B-5E**). Overall, we demonstrated that a high value of the radiomic feature Peri_V_DN predicted poor prognosis for TNBC and more aggressive tumor characteristics.

Integrative Analysis Elucidated Metabolic Reprogramming in High Peri_V_DN Patients

We further investigated the molecular characteristics associated with Peri_V_DN. Using paired transcriptomics (n = 167) (7) and metabolomics (n = 138) data, metabolite abundance and gene

expression were compared between the Peri_V_DN groups (**Supplementary Figure 6A-6C and Supplementary Table 5-7**). Differentially abundant polar metabolites mainly consisted of lipids. Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome-based GSEA demonstrated similar results (FDR < 0.1) that high Peri_V_DN was significantly associated with aberrant metabolism and suppressed immune-related pathways (**Figure 5A, Supplementary Table 8**).

Previous differentially abundant metabolites and pathway enrichment analyses revealed that metabolic reprogramming was related to high Peri_V_DN. On this basis, we performed DA score analysis between the Peri_V_DN groups. Among 53 pathways in which more than three metabolites were annotated, 21 were upregulated and four were downregulated in high Peri_V_DN patients (**Supplementary Table 9**). Three pathways were upregulated with DA scores no less than 0.25, namely, fatty acid biosynthesis, primary bile acid biosynthesis and glutathione metabolism (**Figure 5B**). This result was consistent with that of a previous analysis and further highlighted fatty acid metabolism reprogramming in high Peri_V_DN. We conducted a transcriptomic-metabolomic integrative analysis to depict a more meticulous fatty acid metabolism alteration in this population and to reveal pivotal metabolic targets. Metabolic flux analysis of fatty acid metabolism demonstrated that the initial step of fatty acid synthesis was significantly upregulated, suggesting that *FASN* played a crucial role in fatty acid synthesis rewiring (**Figure 5C**). Acyl carnitines were also highly enriched in high Peri_V_DN and thus sustained an active β-oxidation process. Taken together, these results demonstrated that vigorous *de novo* fatty acid synthesis was closely related to a high Peri_V_DN phenotype.

Distinct Tumor Microenvironments in the Different Peri_V_DN Groups

The landscape of the TNBC microenvironment has been delineated, and distinct "hot tumors" and "cold tumors" have been described (22-24). In this study, we estimated the cell subset composition of the tumor microenvironment by a published gene signature using transcriptomic data (22). We identified 24 different cell types, including immune and stromal cells. RNA-based immune cell deconvolution signatures revealed a major difference in the microenvironment between tumors with high and low peritumor heterogeneity (**Figure 6A-6B**). Low Peri_V_DN was characterized by a stronger signature for CD8+ T cells, naïve CD4+ T cells, γδ T cells, activated NK cells, M1 macrophages and regulatory T cells. A comparison of the estimated cell number between the tumor sites and paired adjacent normal sites also suggested that tumors with low peritumor heterogeneity were composed of more immune cells, both immune-promoting and immune-inhibitory cells (**Supplementary Figure 7A-7B**). Cytolytic activity, which inferred the activity of effector immune cells (25), was lower in high Peri_V_DN cases (*P* = 0.01, **Figure 6C**). These results confirmed that high Peri_V_DN was associated with a suppressed immune response.

Moreover, we investigated the possible immune escape mechanisms of both types of tumors. In addition to the previously described enrichment of regulatory T cells in low Peri_V_DN cases, another inhibitory immune cell type, myeloid-derived suppressor cells (MDSCs), also had a relatively higher abundance in low Peri_V_DN cases (*P* = 0.05, **Figure 6D**) (26). The expression levels of a wide range of immune coinhibitors and costimulators, including multiple immune checkpoints, were investigated, and a

more inhibitory immune context was found in low Peri_V_DN cases (**Figure 6E**). Overall, the delineation of the tumor microenvironment implied that the low Peri_V_DN group tended to have hot tumors and might escape immune surveillance by higher inhibitory immune cell infiltration and stronger immune checkpoint molecule expression.

Furthermore, the comparison of the two common innate immunity sensing pathways, cGAS-STING (27) and NLRP3 inflammasome (28), demonstrated weaker immunity activation in high Peri_V_DN cases (P = 0.03 and P = 0.02, respectively, **Figure 6F**). We also analyzed tumor immunogenicity through MHC molecule expression comparison. Reduced expression of MHC molecules in the high Peri_V_DN group is demonstrated in **Figure 6G**. In summary, the high Peri_V_DN group exhibited a cold tumor phenotype, and its potential escape mechanisms included a reduction in innate immune sensing and rejection of immune infiltration.

Discussion

Recent studies have revealed evident tumor heterogeneity of TNBC, and precision treatment based on molecular profiling has achieved preliminary progress (5, 7, 29, 30). These promising results encouraged molecular subtyping and genomic sequencing for TNBC in clinical practice, which is traditionally conducted by invasive biopsies. Herein, we developed a non-invasive radiomic approach for the identification and molecular classification of TNBC. In addition, we identified prognostic radiomic features with underlying biological properties. These results demonstrated the potential role of a surrogate radiomic approach in distinguishing TNBC patients and further differentiating TNBC into different subtypes and clinical outcomes.

In this study, we first investigated the value of radiomics to distinguish TNBC from other subtypes of breast cancer. We concluded that the non-invasive radiomic approach could identify TNBC with an AUC of 0.922 in the FUSCC cohort and 0.723 in the IPMCH cohort. This conclusion not only informed us of the potential of an MRI-based radiomics approach to identify TNBC with high efficacy but also warned us that the generalization of radiomic models into medical images acquired from multiple medical centers and MRI machines remained an important issue to handle. Other studies also explored the value of radiomics to classify TNBC. Calstaldo and colleagues (31) found that radiomic signatures could identify TNBC with an AUC of 0.91 in MRI images derived from 91 patients, while Leithner and colleagues (32) identified TNBC with an accuracy of 0.736 using a radiomic approach in a cohort consisting of another 91 patients. These results together demonstrated that MRI-based radiomic signatures could classify TNBC with high accuracy and were consistent with the results of our study.

Previous studies have shown that peritumoral heterogeneity can be used to predict the clinical outcomes of several types of cancer (33-35). Herein, we found that Peri_V_DN, which represents the variance in peritumoral heterogeneity on MRI sequences, was associated with adverse clinical outcomes. This conclusion pointed out the unique significance of the peripheral tumor phenotype. Furthermore, to decipher the molecular mechanisms underlying the distinct radiomic patterns, the connection between a

radiomic phenotype and multiomics data needs to be determined, rather than a single driver mutation (11, 36, 37). Several studies have made efforts to explore the association of radiomic features and transcriptomic data (18, 38). Lee et al (16) found that a four-feature radiomic signature could predict the clinical outcomes of pathological T1 renal cell carcinoma and was associated with the abundance of certain immune cell types. This was consistent with our findings, but the exact mechanisms of forming an immunosuppressive microenvironment in tumors with low peritumoral heterogeneity were not explored in this article.

Our study has several limitations. First, more refined models are needed to further improve the prediction efficacy of the radiomic signatures, especially for predicting TNBC molecular subtypes. Second, most patients were recruited from a single institution, and the sample size of the independent external validation cohort was limited (n = 54). Third, the biological characteristics associated with the Peri_V_DN feature were subjected to the nature of exploratory analysis.

In conclusion, we presented a radiomic dataset originating from the largest breast cancer radiomics cohort to date. The radiomics approach showed promising efficacy in identifying TNBC and predicting TNBC molecular subtypes in a non-invasive way. Additionally, peritumoral heterogeneity quantified by radiomics stratified patient outcomes and represented distinct tumor metabolism and immune response patterns. These results demonstrated the potential application of radiomics analysis in the analysis of tumor heterogeneity and the clinical management of TNBC.

Material And Methods

Patient Cohorts and Data Sets

We retrospectively recruited patients diagnosed with malignant breast cancer whose baseline breast CE-MRI images were suitable for radiomics analysis. A total of 860 Chinese patients who were treated at Fudan University Shanghai Cancer Center (FUSCC) from 1 August 2009 to 31 May 2015 were enrolled according to the following defined criteria: 1) female patients diagnosed with unilateral invasive ductal carcinoma with known ER, PR and HER2 phenotypes (39); 2) no evidence of distant metastasis at diagnosis.

To further explore whether radiomics could distinguish different TNBC molecular subtypes, 202 TNBC patients was selected from the FUSCC breast cancer radiomics cohort. These patients were also included in our TNBC multi-omics dataset, which has been published recently (7). Among the 202 patients eligible for the radiomics analysis, transcriptomic sequencing (n = 167), metabolomics (n = 138), hematoxylin and eosin (H.E.)-stained sections with IHC staining (n = 56) for the expression of AR, CD8A, FOXC1 and DCLK1, and clinical follow-up data were also available. The studies were conducted in accordance with the Declaration of Helsinki. All analyses were approved by the independent ethics committee/institutional review board of Fudan University Shanghai Cancer Center, and written informed consent was obtained from each patient.

To evaluate the ability of radiomic signatures to generalize to additional populations, we collected CE-MRI images and ER, PR and HER2 phenotype information from 54 patients diagnosed with malignant breast cancer at the International Peace Maternal and Children Hospital (IPMCH). These patients were recruited from 1 January 2013 to 31 June 2019. Data collection was under approval of the IPMCH institutional review boards.

CE-MRI and Regions of Interest (ROI) Delineation

All patients underwent breast MR examination before biopsy, and CE-MRI images were used for radiomic analysis in this study. The imaging parameters are listed in **Supplementary Table 10**.

ROIs were delineated semiautomatically on the peak enhanced phase of CE-MRI by 3D slicer software (https://www.slicer.org/). ROIs were placed on all slices that contained the whole tumor or the largest lesion (in the case of multicentric or multifocal tumors). Two radiologists (C.Y. and D.D.Z. with 9 and 4 years of experience in breast MRI, respectively) were blinded to the pathological and biochemical findings of each patient and were chiefly responsible for the evaluation of the ROIs. The inter- and intraobserver reproducibility of the ROIs and radiomic feature extraction were initially analyzed with the CE-MRI data of 60 randomly selected patients in a blinded fashion by two radiologists. To ensure reproducibility, one radiologist repeated the ROI drawing twice with an interval of at least 1 month and generated radiomic features following the same procedure. Intraclass correlation coefficients (ICCs) were utilized to evaluate the intra- and interobserver agreement in terms of feature extraction. Inter- and intraobserver reproducibility and radiomic feature extraction achieved substantial agreement with ICC > 0.6 both among the ROIs from the two radiologists and between the ROIs from the same radiologist (40).

Furthermore, the peritumoral area was obtained by expanding the tumor outward with a 2-pixel width and subtracting the tumor area, while the intratumoral area was obtained by shrinking with a 2-pixel width. In addition, tumor and peritumoral regions were integrated as another region. In total, four sets of ROIs, including tumor, peritumor, intratumor and tumor-peritumor regions, were used in the radiomics feature extraction.

Radiomics Feature Extraction

This study extracted the radiomic features of the contrast-enhanced phase, including shape features, first-order features, text features and sequential features, using the PyRadiomics package (41). Fourteen shape-based characteristics, which describe the difference in shape between different types of tumors, were calculated using the first postcontrast phase of CE-MRI. Ninety-one first-order features were calculated from the four phases individually, and 364 features were acquired in total; these features describe the distribution of voxel intensities. Textural features were calculated based on five textural matrices to describe the radiological pattern of the ROI. Moreover, sequential features were calculated to consider time dimension information. The extracted sequential features were mainly composed of two parts: (a) the feature discrepancy between four different phases on the timeline including 364 features

and (b) the mean, variance, kurtosis and skewness of the time curve extracted from each individual patient, including ninety-one in each group. As described in the ROI delineation, feature extraction was performed in the tumoral, peritumoral, intratumoral and tumor-peritumoral regions.

Feature Selection and Radiomics Model Building

The least absolute shrinkage and selection operator (LASSO) method was used to select the most useful predictive features from the training cohort (glmnet R package) (42). Tuning parameter (λ) selection in the LASSO model used 9-fold cross-validation. Radiomics scores were calculated for each patient via three different methods: 1) multivariate linear regression (glm R package), 2) support vector machine (SVM; e1071 R package), and 3) deep neural network (DNN; sklearn Python package). The ability to predict TNBC molecular subtypes was assessed by the area under the receiver operating characteristic curve (AUC) of the receiver operator characteristic (ROC) curve via the pROC R package (43). Confidence intervals of AUCs were calculated using the Delong method.

Radiomics Model Validation

The radiomics prediction models were validated internally and externally. First, the trained classifiers were assessed by cross-validation via the glmnet R package (42). Then, the trained classifiers were further tested in the validation cohort in terms of the AUC and its confidence intervals of the ROC curve.

Collection and Analysis of Metabolomics and Lipidomics Data

Samples in our multiomics TNBC cohort with adequate tissues for polar metabolites and lipids were collected. In total, 138 TNBC samples were selected for further metabolomics and lipidomics analysis. Acetonitrile: methanol: water = 2: 2: 1 solution and MTBE: MeOH= 5: 1 solution were applied to extract polar metabolites and lipids, respectively. An equal volume (10 μL) of each sample was mixed for quality control sample preparation. A BEH amide column (2.1 * 100 mm, 1.7 μm, Waters) or Kinetex C18 column (2.1 * 100 mm, 1.7 μm, Phenomen) coupled with a Triple TOF 6600 mass spectrometer or AB triple TOF 5600 mass spectrometer was deployed to conduct LC-MS/MS experiments for polar metabolite and lipid detection. Detailed information on sample processing, metabolomics and lipidomics data generation is contained in the supplementary materials and methods.

Analysis of Differentially Abundant Metabolites and Differentially Expressed Genes

The differential abundance of metabolites was calculated by performing Mann-Whitney U tests for all detected metabolites. The differential expression of genes was determined via the limma R package (44). Subsequent gene set enrichment analysis (GSEA) was performed using the clusterProfiler R package (45). The differential expression analysis outputs of limma were used to generate the ranked list file. One thousand total permutations were used.

Differential Abundance (DA) Score

The DA score was calculated first by determining which metabolites were significantly increased/decreased in abundance, as described above. Then, the DA score was defined as follows:

DA = (Number of metabolites increased - Number of metabolites decreased)/Number of measured metabolites in that pathway (46)

Thus, the DA score ranges from -1 to 1. A score of -1 indicates that all metabolites in a pathway decreased, while a score of 1 indicates that all metabolites increased in abundance.

Calculation of Microenvironment Cell Abundance

A signature containing 364 genes representing 24 microenvironment cell types was obtained from one published high-quality paper (22). This signature modified the CIBERSORT and MCP-Counter signatures and represented a more comprehensive landscape of the TNBC microenvironment. Subsequently, we used single-sample gene set enrichment analysis (ssGSEA, "GSVA" function in GSVA R package) (47) to calculate the abundance of each cell subset in each sample with expression data. Additionally, we referred to the xCell signature to further validate the composition of the tumor microenvironment (48).

Statistical Analysis

Student's t test, Wilcoxon's test and Kruskal-Wallis test were used to compare continuous variables. Prior to the comparisons, the normality of the distributions was tested with the Shapiro-Wilk test. Pearson's chi-square test and Fisher's exact test were employed for the comparison of unordered categorical variables. To explore the association between radiomics features and survival, Kaplan-Meier analysis and a Cox proportional hazards model were employed in the training and validation cohorts. The comparison of survival between groups was conducted via the log rank test. All the tests were two sided, and $P < 0.05$ was regarded as indicating significance unless otherwise stated. In multiple hypothesis testing, false discovery rate (FDR) correction was used to decrease false positive rates. All statistical analyses were performed with R software (version 3.6.1, http://www.R-project.org).

Declarations

Conception and design: Zhi-Ming Shao, Yi-Zhou Jiang and Ya-Jia Gu.

Development of methodology: Lin Jiang, Chao You, Yi Xiao and Ren-Cheng Zheng.

Acquisition of data: Chao You, Ren-Cheng Zheng, Bing-Qing Xia, He Wang and Dan-Dan Zhang.

Analysis and interpretation of data: Lin Jiang and Yi Xiao.

Writing, review, and/or revision of the manuscript: Lin Jiang, Chao You, Yi Xiao and Yi-Zhou Jiang.

Study supervision: Zhi-Ming Shao, Yi-Zhou Jiang and Ya-Jia Gu.

References

1.   Waks AG, and Winer EP. Breast Cancer Treatment: A Review. *JAMA.* 2019;321(3):288-300.

2.   Criscitiello C, Azim HA, Jr., Schouten PC, Linn SC, and Sotiriou C. Understanding the biology of triple-negative breast cancer. *Ann Oncol.* 2012;23 Suppl 6:vi13-8.

3.   Denkert C, Liedtke C, Tutt A, and von Minckwitz G. Molecular alterations in triple-negative breast cancer—the road to new treatment strategies. *The Lancet.* 2017;389(10087):2430-42.

4.   Metzger-Filho O, Tutt A, de Azambuja E, Saini KS, Viale G, Loi S, et al. Dissecting the heterogeneity of triple-negative breast cancer. *J Clin Oncol.* 2012;30(15):1879-87.

5.   Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest.* 2011;121(7):2750-67.

6.   Burstein MD, Tsimelzon A, Poage GM, Covington KR, Contreras A, Fuqua SA, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res.* 2015;21(7):1688-98.

7.   Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell.* 2019;35(3):428-40 e5.

8.   Dagogo-Jack I, and Shaw AT. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol.* 2018;15(2):81-94.

9.   Marusyk A, Janiszewska M, and Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell.* 2020;37(4):471-84.

10.   Zhao S, Zuo WJ, Shao ZM, and Jiang YZ. Molecular subtypes and precision treatment of triple-negative breast cancer. *Ann Transl Med.* 2020;8(7):499.

11.   Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res.* 2017;77(14):3922-30.

12.   Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* 2012;48(4):441-6.

13.   Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin.* 2019;69(2):127-57.

14.   Jansen RW, van Amstel P, Martens RM, Kooi IE, Wesseling P, de Langen AJ, et al. Non-invasive tumor genotyping using radiogenomic biomarkers, a systematic review and oncology-wide pathway analysis. *Oncotarget.* 2018;9(28):20134-55.

15.   Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol.* 2016;34(18):2157-64.

16.   Lee HW, Cho HH, Joung JG, Jeon HG, Jeong BC, Jeon SS, et al. Integrative Radiogenomics Approach for Risk Assessment of Post-Operative Metastasis in Pathological T1 Renal Cell Carcinoma: A Pilot Retrospective Cohort Study. *Cancers (Basel).* 2020;12(4).

17.   Fischer S, Tahoun M, Klaan B, Thierfelder KM, Weber MA, Krause BJ, et al. A Radiogenomic Approach for Decoding Molecular Mechanisms Underlying Tumor Progression in Prostate Cancer. *Cancers (Basel).* 2019;11(9).

18.   Lin P, Wen DY, Chen L, Li X, Li SH, Yan HB, et al. A radiogenomics signature for predicting the clinical outcome of bladder urothelial carcinoma. *Eur Radiol.* 2020;30(1):547-57.

19.   Kim JH, Ko ES, Lim Y, Lee KS, Han BK, Ko EY, et al. Breast Cancer Heterogeneity: MR Imaging Texture Analysis and Survival Outcomes. *Radiology.* 2017;282(3):665-75.

20.   Wu J, Gong G, Cui Y, and Li R. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imaging.* 2016;44(5):1107-15.

21.   Zhao S, Ma D, Xiao Y, Li XM, Ma JL, Zhang H, et al. Molecular Subtyping of Triple-Negative Breast Cancers by Immunohistochemistry: Molecular Basis and Clinical Relevance. *Oncologist.* 2020.

22.   Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. *Clin Cancer Res.* 2019.

23.   O'Meara T, Safonov A, Casadevall D, Qing T, Silber A, Killelea B, et al. Immune microenvironment of triple-negative breast cancer in African-American and Caucasian women. *Breast Cancer Res Treat.* 2019;175(1):247-59.

24.   Keren L, Bosse M, Marquez D, Angoshtari R, Jain S, Varma S, et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell.* 2018;174(6):1373-87.e19.

25.   Rooney Michael S, Shukla Sachet A, Wu Catherine J, Getz G, and Hacohen N. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell.* 2015;160(1-2):48-61.

26.   Angelova M, Charoentong P, Hackl H, Fischer ML, Snajder R, Krogsdam AM, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* 2015;16:64.

27.   Woo SR, Fuertes MB, Corrales L, Spranger S, Furdyna MJ, Leung MY, et al. STING-dependent cytosolic DNA sensing mediates innate immune recognition of immunogenic tumors. *Immunity.* 2014;41(5):830-42.

28.   Ghiringhelli F, Apetoh L, Tesniere A, Aymeric L, Ma Y, Ortiz C, et al. Activation of the NLRP3 inflammasome in dendritic cells induces IL-1beta-dependent adaptive immunity against tumors. *Nat Med.* 2009;15(10):1170-8.

29.   Jiang YZ, Liu Y, Xiao Y, Hu X, Jiang L, Zuo WJ, et al. Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: the FUTURE trial. *Cell Res.* 2020.

30.   Garrido-Castro AC, Lin NU, and Polyak K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov.* 2019;9(2):176-98.

31.   Castaldo R, Pane K, Nicolai E, Salvatore M, and Franzese M. The Impact of Normalization Approaches to Automatically Detect Radiogenomic Phenotypes Characterizing Breast Cancer Receptors Status. *Cancers (Basel).* 2020;12(2).

32.   Leithner D, Horvat JV, Marino MA, Bernard-Davila B, Jochelson MS, Ochoa-Albiztegui RE, et al. Radiomic signatures with contrast-enhanced magnetic resonance imaging for the assessment of breast cancer receptor status and molecular subtypes: initial results. *Breast Cancer Res.* 2019;21(1):106.

33.   Dou TH, Coroller TP, van Griethuysen JJM, Mak RH, and Aerts H. Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PLoS One.* 2018;13(11):e0206108.

34.   Prasanna P, Patel J, Partovi S, Madabhushi A, and Tiwari P. Radiomic features from the peritumoral brain parenchyma on treatment-naïve multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: Preliminary findings. *Eur Radiol.* 2017;27(10):4188-97.

35.   Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res.* 2017;19(1):57.

36.   Gierach GL, Li H, Loud JT, Greene MH, Chow CK, Lan L, et al. Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study. *Breast Cancer Res.* 2014;16(4):424.

37.   Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, et al. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer.* 2019;132:28-35.

38.   Yeh AC, Li H, Zhu Y, Zhang J, Khramtsova G, Drukker K, et al. Radiogenomics of breast cancer using dynamic contrast enhanced MRI and gene expression profiling. *Cancer Imaging.* 2019;19(1):48.

39.   Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, 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(16):2784-95.

40.   Landis JR, and Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-74.

41.   van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017;77(21):e104-e7.

42.   Friedman J, Hastie T, and Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.

43.   Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011;12:77.

44.   Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47.

45.   Yu G, Wang LG, Han Y, and He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics.* 2012;16(5):284-7.

46.   Hakimi AA, Reznik E, Lee CH, Creighton CJ, Brannon AR, Luna A, et al. An Integrated Metabolic Atlas of Clear Cell Renal Cell Carcinoma. *Cancer Cell.* 2016;29(1):104-16.

47.   Hänzelmann S, Castelo R, and Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.

48.   Aran D, Hu Z, and Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.

49.   Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013;4:2612.

Tables

Table 1 Multivariate Cox proportional hazard models for RFS and OS

Variables		RFS		OS	
		HR (95% CI)	P	HR (95% CI)	P
T stage	T1		Ref		
	T2	0.58 (0.27-1.25)	0.17	0.70 (0.24-2.02)	0.51
	T3/T4	0.95 (0.24-3.76)	0.94	1.69 (0.29-9.75)	0.59
N stage	pN0		Ref		
	pN1	2.03 (0.81-5.07)	0.13	2.73 (0.77-9.66)	0.12
	pN2	5.02 (1.88-13.43)	0.001	5.87 (1.66-20.76)	0.006
	pN3	7.62 (2.84-20.43)	5.38e-05	4.85 (1.06-22.24)	0.04
TNBC subtype	BLIS		Ref		
	IM	1.12 (0.35-3.52)	0.85	0.67 (0.13-3.49)	0.63
	MES	0.93 (0.29-2.96)	0.90	0.82 (0.19-3.46)	0.78
	LAR	0.94 (0.34-2.57)	0.90	0.62 (0.16-2.30)	0.47
Peri_V_DN	High		Ref		
	Low	0.41 (0.18-0.95)	0.04	0.15 (0.03-0.70)	0.02

Abbreviations: OS, overall survival; Peri_V_DN, variance of dependence non-uniformity extracted of peritumoral regions; RFS, recurrence-free survival; TNBC, triple-negative breast cancer.
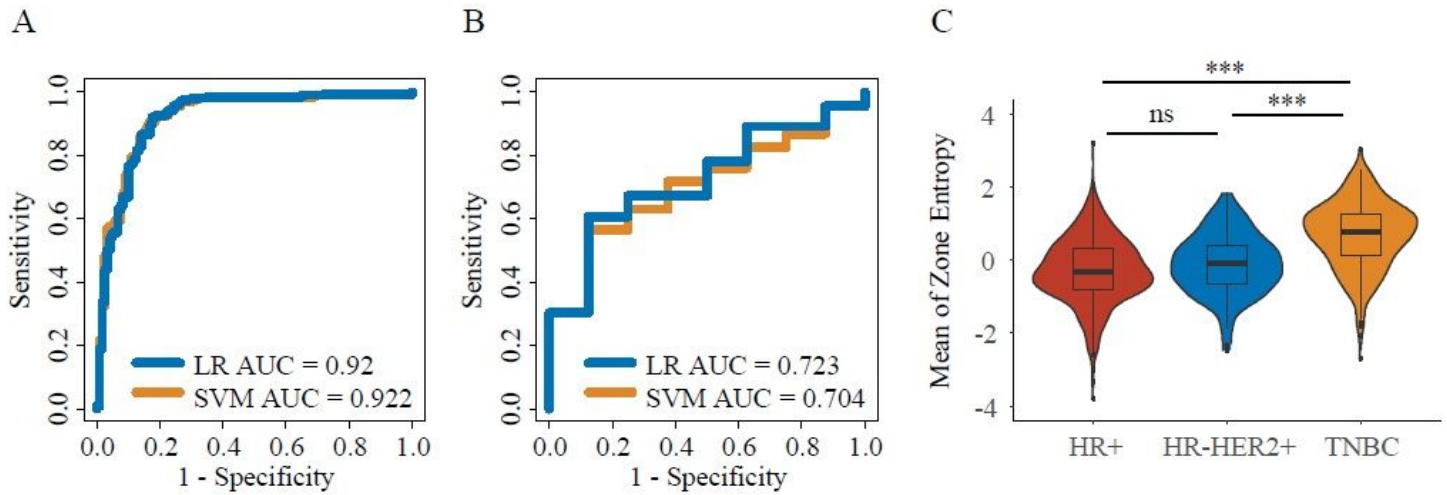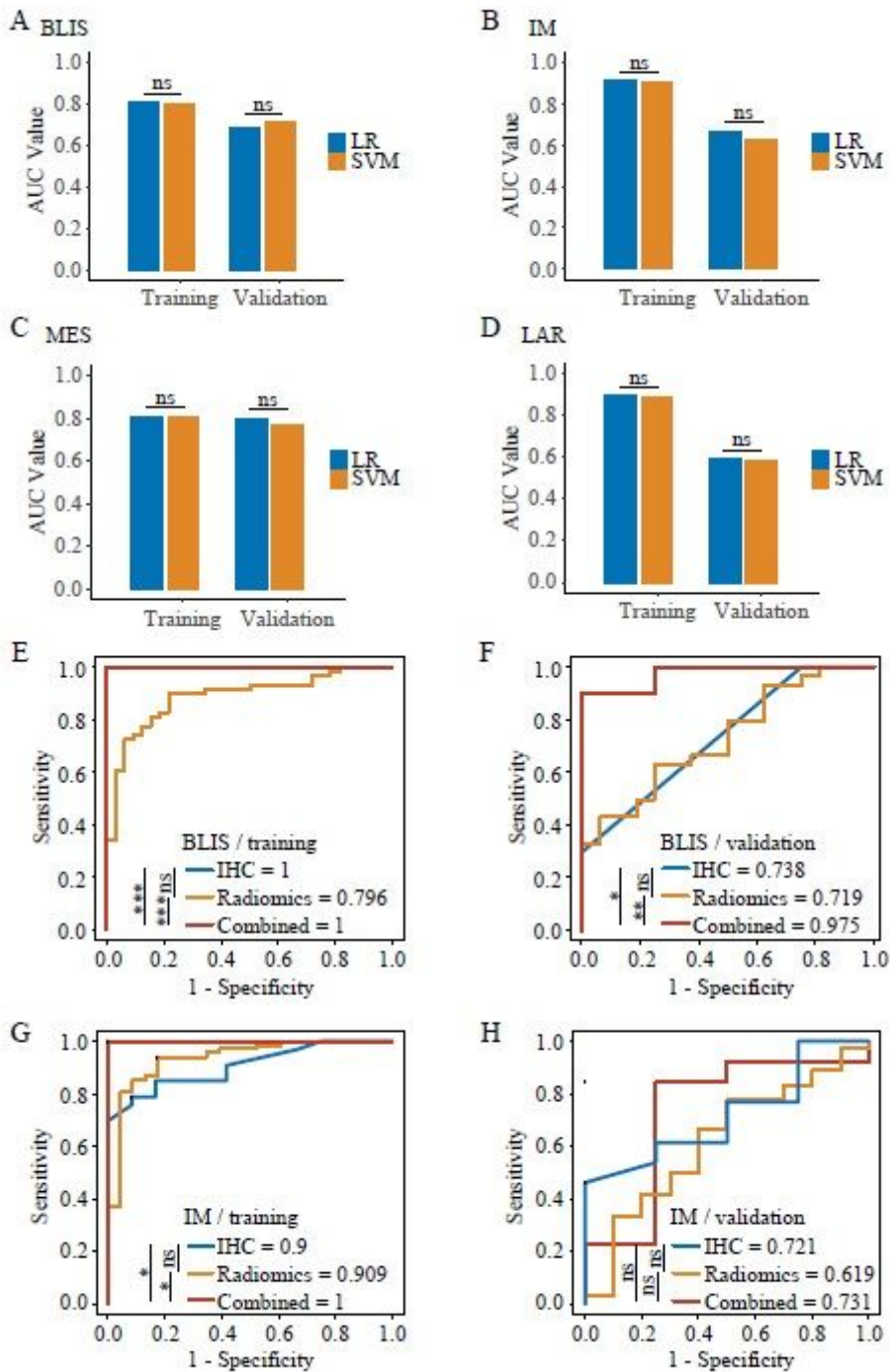
Figures
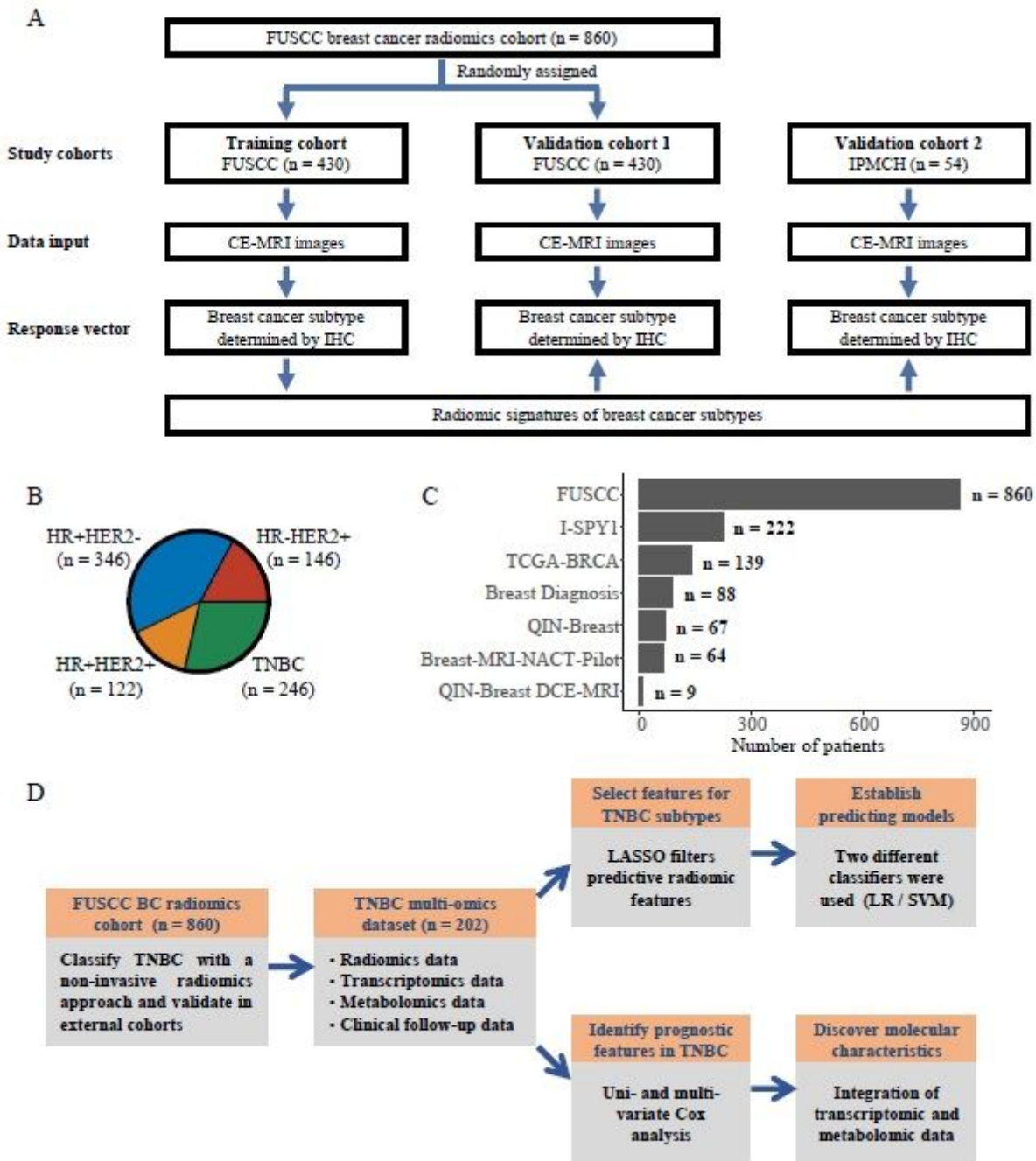
Figure 1

Workflow of the radiogenomics analysis and data description. (A) Study cohorts used in this study. (B) Composition of patients enrolled in the FUSCC breast cancer radiomics cohort according to receptor status. (C) Sample sizes of the FUSCC breast cancer radiomics cohort and other cohorts based on MRI images from the TCIA database. (D) Analysis plan of integrative radiogenomics analysis. Abbreviations: CE-MRI, contrast-enhanced magnetic resonance imaging; FUSCC, Fudan University Shanghai Cancer Center; HR, hormone receptor; HER-2, human epidermal growth factor receptor-2; IHC, immunohistochemistry; IPMCH, International Peace Maternity and Children Hospital; LASSO, least

absolute shrinkage and selection operator; LR, logistic regression; SVM, support vector machine; TNBC, triple-negative breast cancer;
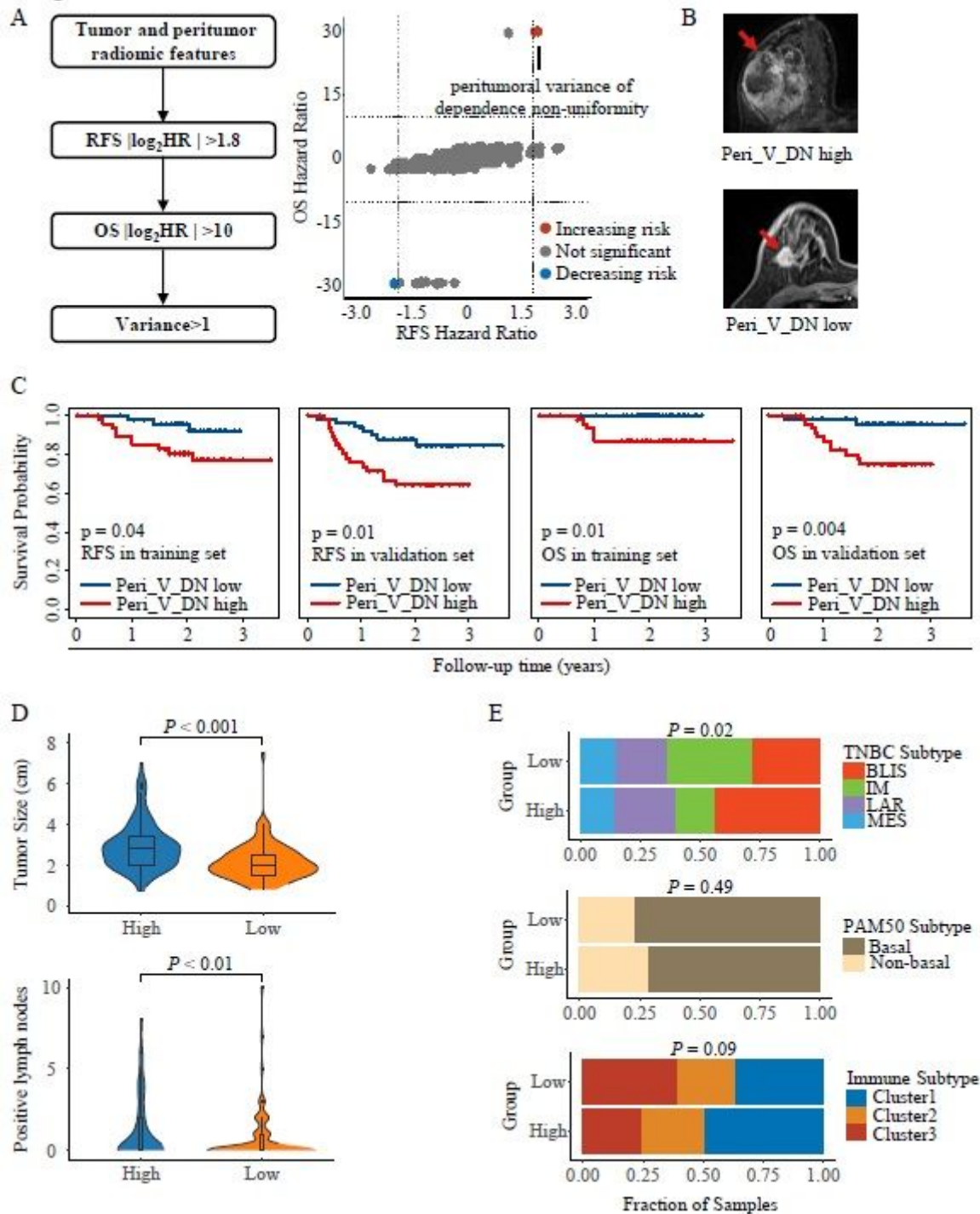

Figure 2

Classification of TNBC molecular subtypes via a radiomic approach with machine learning method. (A) AUCs of the radiomic signature for distinguishing TNBC from other breast cancer subtypes in validation cohort 1. (B) AUCs of the radiomic signature for distinguishing TNBC from other breast cancer subtypes in validation cohort 2. (C) Mean of tumor zone entropy, which is a radiomic feature that evaluates tumor heterogeneity, is significantly higher in TNBC than in other breast cancer subtypes. *** $P < 0.001$; ** $0.001 < P < 0.01$; * $0.01 < P < 0.05$; ns $P > 0.05$. Abbreviations: AUC, area under the receiver operating characteristic curve; HR, hormone receptor; HER-2, human epidermal growth factor receptor-2; LR, logistic regression; SVM, support vector machine; TNBC, triple-negative breast cancer.

Figure 3

Efficacy of predicting TNBC molecular subtypes using radiomic and IHC data with machine learning method. (A-D) AUCs of the radiomic signatures for predicting BLIS (A), IM (B), MES (C) and LAR (D) subtypes. (E-H) Comparison of the combined model, individual radiomic model and IHC model for predicting BLIS (E and F) and IM (G and H) subtypes. The sample size of each prediction model was 90 in the training cohort and 46 in the validation cohort. Abbreviations: BLIS, basal-like immune-suppressed; DNN, deep neural network; IHC, immunohistochemistry; IM, immunomodulatory; LAR, luminal androgen

receptor; LR, logistic regression; MES, mesenchymal-like; SVM, support vector machine; TNBC, triple-negative breast cancer.

Figure 4

Identification of the prognostic feature Peri_V_DN and its clinicopathological associations. (A) Criteria of prognostic feature selection (left) and hazard ratios for RFS and OS of the radiomic features (right). (B) Breast CE-MRI images from one patient with high Peri_V_DN and one patient with low Peri_V_DN. (C) Kaplan-Meier plots show the prognostic power of Peri_V_DN for RFS and OS. (D) Tumor size and pathologically confirmed metastatic lymph nodes between the high and low Peri_V_DN groups. (E)

Distribution of the TNBC mRNA subtypes, PAM50 subtypes and TNBC microenvironment clusters between the high and low Peri_V_DN groups. Abbreviations: HR, hazard ratio; OS, overall survival; Peri_V_DN, peritumoral variance in dependence nonuniformity of peritumoral regions; RFS, recurrence-free survival.
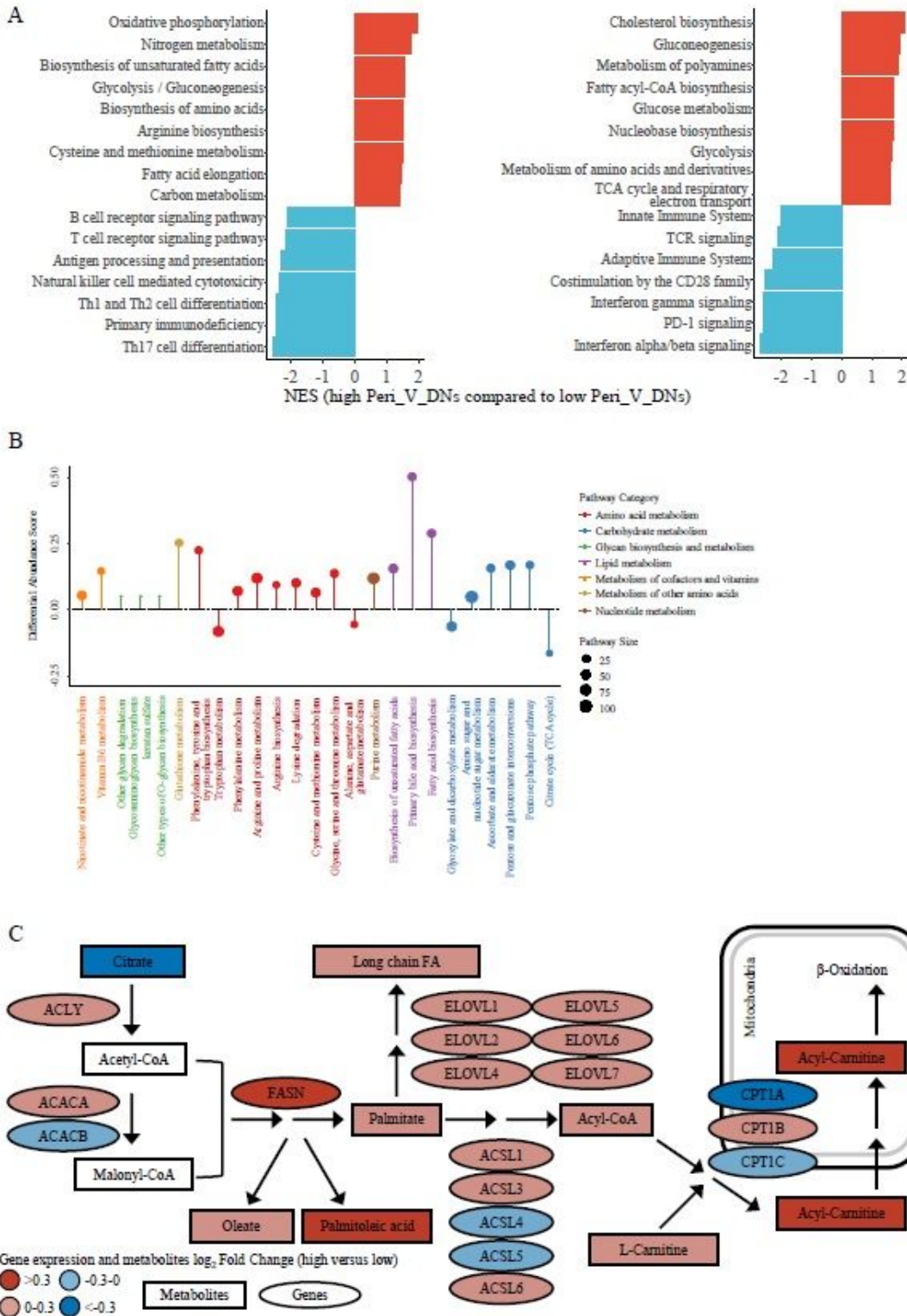

Figure 5

Identification of differentially expressed pathways and metabolic flux analysis. (A) GSEA revealed that metabolic pathways were upregulated and immune-related pathways were downregulated in the high Peri_V_DN group compared to the low Peri_V_DN group. (B) The differential abundance score illustrated major differential metabolic pathways utilizing metabolomic data. (C) Metabolic flux analysis combining transcriptomic and metabolomic data showed reprogramming of the fatty acid biosynthesis pathway between the Peri_V_DN groups. (D) GSVA revealed differences in multiple lipid metabolism pathways. In total, 167 samples with transcriptomics data and 138 samples with metabolomics data were included for analysis. Abbreviations: FA, fatty acid; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; NES, normalized enrichment score; Peri_V_DN, peritumoral variance in dependence nonuniformity of peritumoral regions; ssGSEA, single-sample gene set enrichment analysis.
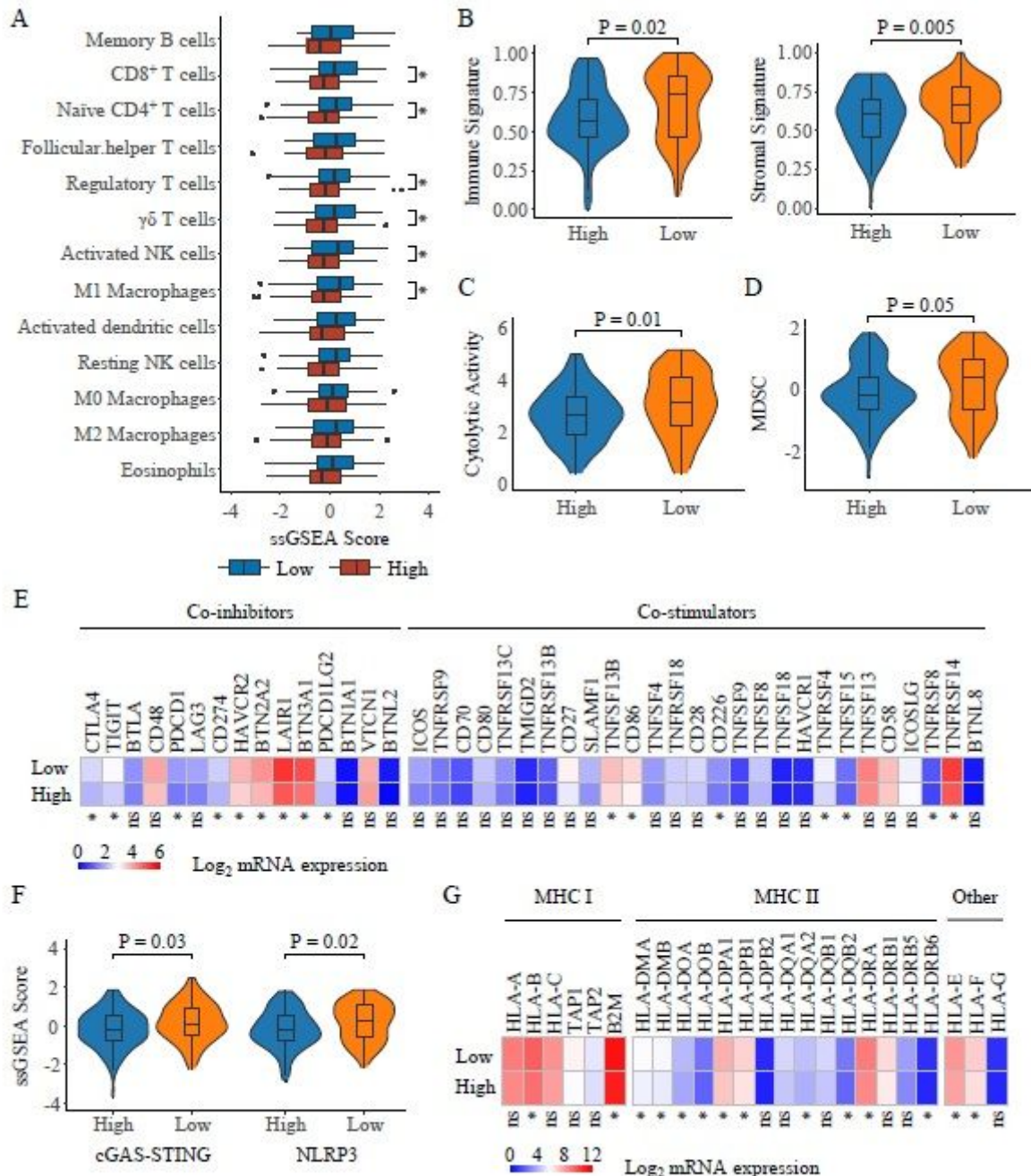

Figure 6

Landscape of the tumor microenvironment of the Peri_V_DN groups and distinct escape mechanisms. (A) Differences in the abundance of immune cell types revealed an immunosuppressive tumor microenvironment in the high Peri_V_DN group compared with the low Peri_V_DN group. (B) The immune signature (left) and stromal signature (right) inferred by ESTIMATE (49) supported an immunosuppressive tumor microenvironment in the high Peri_V_DN group. (C) Comparison of cytolytic activity showed higher effector immune cell activity in the low Peri_V_DN group. (D) Comparison of the proportion of MDSCs between the Peri_V_DN groups showed higher MDSC infiltration in the low Peri_V_DN group. (E) Multiple immune coinhibitors and costimulators had higher expression levels in the low Peri_V_DN group. (F) Two common innate immunity sensing pathways, cGAS-STING and the NLRP3 inflammasome, were downregulated in the high Peri_V_DN group. (G) Comparison of the log2-fold changes in mRNA expression of the MHC molecules between the Peri_V_DN groups. In total, 167 samples with transcriptomics data were included for analysis. Abbreviations: GSVA, gene set variation analysis; MDSC, myeloid-derived suppressor cell; Peri_V_DN, peritumoral variance in dependence nonuniformity of peritumoral regions.

Supplementary Files

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

- 03SupplementaryFigures.pdf
- 04SupplementaryMaterialsandMethods.docx
- 05SupplementaryTables.xlsx