

RESEARCH

Open Access



# Radiomics features for assessing tumor-infiltrating lymphocytes correlate with molecular traits of triple-negative breast cancer

Guan-Hua Su<sup>1,2†</sup>, Yi Xiao<sup>1,2†</sup>, Lin Jiang<sup>1,2</sup>, Ren-Cheng Zheng<sup>3</sup>, He Wang<sup>3</sup>, Yan Chen<sup>4</sup>, Ya-Jia Gu<sup>2,5\*</sup>, Chao You<sup>2,5\*</sup> and Zhi-Ming Shao<sup>1,2</sup>

## Abstract

**Background:** Tumor-infiltrating lymphocytes (TILs) have become a promising biomarker for assessing tumor immune microenvironment and predicting immunotherapy response. However, the assessment of TILs relies on invasive pathological slides.

**Methods:** We retrospectively extracted radiomics features from magnetic resonance imaging (MRI) to develop a radiomic cohort of triple-negative breast cancer (TNBC) (n = 139), among which 116 patients underwent transcriptomic sequencing. This radiomic cohort was randomly divided into the training cohort (n = 98) and validation cohort (n = 41) to develop radiomic signatures to predict the level of TILs through a non-invasive method. Pathologically evaluated TILs in the H&E sections were set as the gold standard. Elastic net and logistic regression were utilized to perform radiomics feature selection and model training, respectively. Transcriptomics was utilized to infer the detailed composition of the tumor microenvironment and to validate the radiomic signatures.

**Results:** We selected three radiomics features to develop a TILs-predicting radiomics model, which performed well in the validation cohort (AUC 0.790, 95% confidence interval (CI) 0.638–0.943). Further investigation with transcriptomics verified that tumors with high TILs predicted by radiomics (Rad-TILs) presented activated immune-related pathways, such as antigen processing and presentation, and immune checkpoints pathways. In addition, a hot immune microenvironment, including upregulated T cell infiltration gene signatures, cytokines, costimulators and major histocompatibility complexes (MHCs), as well as more CD8<sup>+</sup> T cells, follicular helper T cells and memory B cells, was found in high Rad-TILs tumors.

**Conclusions:** Our study demonstrated the feasibility of radiomics model in predicting TILs status and provided a method to make the features interpretable, which will pave the way toward precision medicine for TNBC.

**Keywords:** Radiomics, Tumor-infiltrating lymphocytes (TILs), Triple-negative breast cancer (TNBC), Tumor microenvironment (TME), Magnetic resonance imaging (MRI)

<sup>†</sup>Guan-Hua Su and Yi Xiao are contributed equally to this work

\*Correspondence: [cjr.guyajia@vip.163.com](mailto:cjr.guyajia@vip.163.com); [youchao8888@aliyun.com](mailto:youchao8888@aliyun.com)

<sup>5</sup> Department of Radiology, Fudan University Shanghai Cancer Center, No. 270 Dong'an Road, Shanghai 200032, People's Republic of China  
Full list of author information is available at the end of the article

## Introduction

Triple-negative breast cancer (TNBC) is defined as a breast cancer subtype that lacks expression of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor type 2 (HER2) [1]. Due



to the aggressive biological nature of TNBC and the lack of therapeutic targets, TNBCs are characterized by frequent local recurrence and visceral metastasis [1, 2].

Tumor-infiltrating lymphocytes (TILs) have been used as a biomarker of prognosis and therapeutic response in several cancer types [3, 4]. In breast cancer, TILs are most commonly found in TNBC [5, 6]. In a series of clinical trials and prospective studies, recurrence-free survival (RFS), disease-free survival (DFS) and overall survival (OS) outcomes were positively correlated with the quantity of TILs in TNBC tumors [5–9]. Lymphocyte-predominant breast cancer (LPBC) is considered to be a type of breast cancer that responds better to chemotherapy than non-lymphocyte-predominant breast cancer (non-LPBC) [10–12]. In recent years, immunotherapy, particularly the use of immune checkpoint blockades (ICBs), has produced favorable clinical benefits in patients with both early and advanced TNBC [13–15]. In addition to current biomarkers [programmed cell death-ligand 1 (PD-L1), tumor mutation burden (TMB) and microsatellite instability/deficient mismatch repair (MSI/dMMR)] [16, 17], TILs are expected to become another biomarker for predicting patient response to ICBs. Currently, TILs are evaluated through features exhibited by hematoxylin and eosin (H&E)-stained pathological slides obtained via invasive biopsy [3].

Radiomics is a method for extracting high-throughput features from medical images [18, 19]. These quantitative

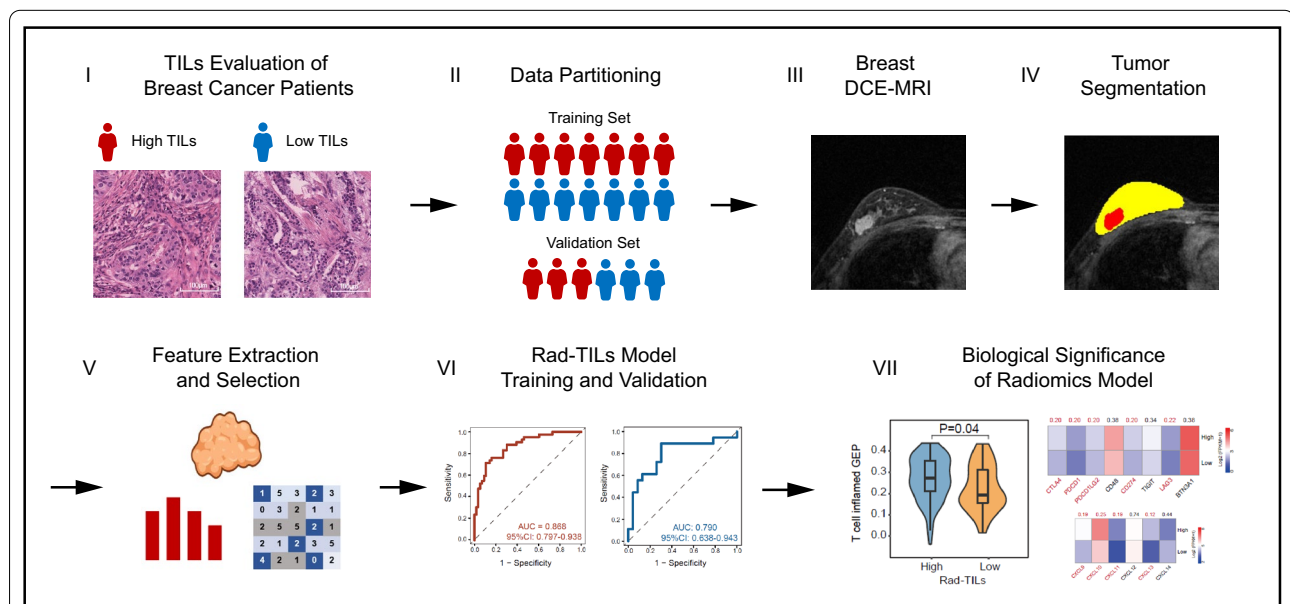
features could be analyzed with data from other observations to reflect the presence of significant genomic events, patients’ response to therapy and prognosis, ultimately contributing to cancer diagnosis and treatment [18–24]. The noninvasive and reproducible nature of radiomics provides us with a favorable approach to predict clinicopathological variables. However, radiomics is limited by its poor interpretability.

In this article, we developed a radiomics signature to infer TILs status noninvasively and investigate the molecular biological significance of the radiomics signature, hoping to overcome the poor interpretability and facilitate the clinical utilization of radiomics for TNBC treatment optimization.

## Methods

### Cohorts and datasets

We retrospectively enrolled 139 triple-negative breast cancer (TNBC) patients treated at the Fudan University Shanghai Cancer Center (FUSCC) from 1 August 2009 to 31 May 2015 with baseline dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) available who were suitable for radiomics analysis. In this TNBC radiomic cohort, transcriptomic data (n = 116) was also available. The framework of this study is presented as Fig. 1. The TNBC radiomics cohort (n = 139) was split into a training cohort (n = 98) and a validation cohort (n = 41) with a 7:3 ratio



**Fig. 1** Schematic of the study. Tumor-infiltrating lymphocytes (TILs) densities were evaluated on H&E slides and were split into high and low TILs based on cut-off of 20%. Study cohort was randomly divided into training and validation cohort at a 7:3 ratio and similar high TILs proportion was kept in training and validation cohort. Regions of interest (ROIs) were segmented from the original breast MRI. Radiomics features were extracted from ROIs and were used to develop a TILs prediction model. Transcriptomics analysis was performed to further illustrate the radiomics model

using a stratified randomization method to keep high and low TILs proportions similar in the two cohorts (Table 1). Quantification of stromal tumor-infiltrating lymphocytes (sTILs), fibrosis and necrosis were evaluated on pathological H&E staining area by two pathologists according to published guidelines [25, 26]. In this study, tumor-infiltrating lymphocytes (TILs) refer to sTILs unless otherwise specified. A percentage of sTILs  $\geq 20\%$  was defined as a high TILs level (Fig. 2A, B).

#### Magnetic resonance imaging (MRI) parameters

All the patients in this cohort underwent MRI with 1.5 T special breast magnetic resonance (Aurora Imaging Technology, Aurora Systems, Inc., Canada) and coils for breast. A series of cross-sectional images were obtained in prone position, including plain scan T2WI (TR 6680 ms, TE 68 ms, slice thickness 3 mm, slice spacing 1 mm), T1WI (TR 5 ms, TE 13 ms, slice thickness 3 mm, slice spacing 1 mm) and dynamic contrast-enhanced T1WI (TR 5 ms, TE 29 ms, slice thickness 1.1 mm, slice spacing 0 mm, FOV 360  $\times$  360 mm). The contrast medium Gd-DTPA (0.2 mmol/kg, flow rate 2.0 ml/s) was injected 90 s after plain scan. Postcontrast images were obtained at 90, 180, 270, and 360 s after injection.

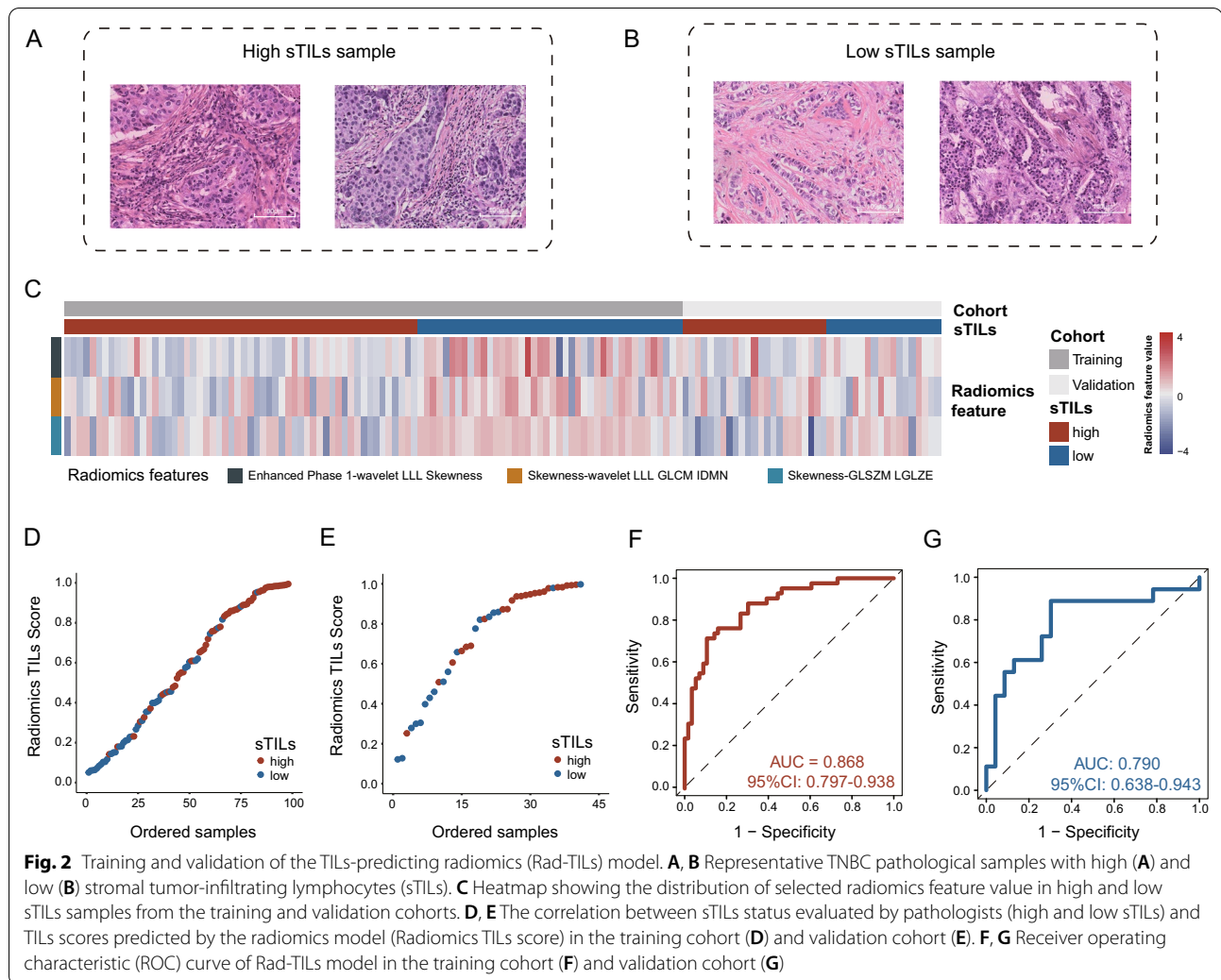
#### Image preprocessing

In this study, the tumor regions of interest (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). To ensure reproducibility, some of the ROIs were initially delineated by two radiologists at FUSCC (C.Y. and D.D.Z. with 9 and 4 years of experience in breast MRI, respectively). The inter- and intra-observer 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. Additionally, one radiologist (C.Y. with 9 years of experience in breast MRI) 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  $>0.75$  both among the ROIs from the two radiologists and between the ROIs from the same radiologist [27]. An ICC greater than 0.6 was considered a marker of satisfactory inter- and intra-observer reproducibility. On the premise of good consistency, whole ROI segmentation

**Table 1** Comparison of the basic information of the training and validation sets

| Clinical characters    | Level   | Training set (n = 98) | Validation set (n = 41) | p value |
|------------------------|---------|-----------------------|-------------------------|---------|
| Age (years, mean [SD]) | –       | 54.28 (10.57)         | 54.61 (12.25)           | 0.872   |
| Menopause (%)          | False   | 31 (32.3)             | 13 (32.5)               | 1       |
|                        | True    | 65 (67.7)             | 27 (67.5)               |         |
| T stage (%)            | 1       | 48 (49.0)             | 14 (34.1)               | 0.206   |
|                        | 2       | 46 (46.9)             | 26 (63.4)               |         |
|                        | 3       | 4 (4.1)               | 1 (2.4)                 |         |
| N stage (%)            | 0       | 56 (58.3)             | 27 (65.9)               | 0.226   |
|                        | 1       | 23 (24.0)             | 12 (29.3)               |         |
|                        | 2       | 12 (12.5)             | 2 (4.9)                 |         |
|                        | 3       | 5 (5.2)               | 0 (0.0)                 |         |
| Stromal TILs (%)       | High    | 56 (57.1)             | 23 (56.1)               | 1       |
|                        | Low     | 42 (42.9)             | 18 (43.9)               |         |
| Ki-67 [mean (SD)]      | –       | 56.99 (21.90)         | 64.27 (26.14)           | 0.094   |
| Molecular subtype (%)  | BLIS    | 29 (29.6)             | 13 (31.7)               | 0.709   |
|                        | IM      | 21 (21.4)             | 9 (22.0)                |         |
|                        | LAR     | 17 (17.3)             | 8 (19.5)                |         |
|                        | MES     | 16 (16.3)             | 3 (7.3)                 |         |
|                        | Unknown | 15 (15.3)             | 8 (19.5)                |         |

SD standard deviation, BLIS basal-like immune-suppressed, IM immunomodulatory, LAR luminal androgen receptor, MES mesenchymal-like



was completed by the more experienced radiologist in each layer of the MRI scan.

All other phases were co-registered into the first postcontrast phase of DCE-MRI through non-linear registration using the symmetric normalization algorithm [28], which was performed in the ANTs toolbox, to eliminate the spatial mismatches caused by motion artifact. Nonparametric nonuniformity normalization (N3) algorithm was applied for bias field correction [29]. Moreover, Z-Score Normalization algorithm was used for data normalization.

#### Radiomics feature extraction

We performed a feature extraction process on DCE-MRI images (four phases) based on the open source Pyradiomics package V3.0, implemented in Python 3.6 [30], including shape features, first-order features, textural features, wavelet domain features and time

domain features. For spatial domain features, 14 shape-based features were common to all phases, which describe the difference in shape between different types of tumors. Eighteen first-order features and 75 textural features were calculated from the four phases individually. First-order features describe the distribution of voxel intensities, and textural features were obtained based on 5 textural matrices to describe the radiological pattern of the ROI, including the gray level co-occurrence matrix (GLCM), gray level dependence matrix (GLDM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM). Moreover, wavelet domain features were extracted for each first order feature and textural feature by applying wavelet filtering to the original images, yielding 8 decompositions per level. In addition, for time domain features, the extracted sequential features were mainly composed of

the mean, variance, kurtosis and skewness of the time-varying curve constructed based on feature values in four phases, for each first order, textural and wavelet domain feature. The specific number of features and the corresponding calculation formulae are described in detail in the Additional file 1.

### Model training and validation

Elastic net regression and logistic regression were utilized to select the most predictive radiomics features from the extracted features and to train machine learning model, respectively. Specifically, in the training cohort, we selected the most predictive radiomics features characterizing TILs levels with elastic net regression [31]. Then, the logistic regression was performed with the selected features to develop a TILs prediction signature referred to as Rad-TILs. The probability of high TILs predicted by radiomics model ( $p$ ) was generated by the following formula:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

In this study,  $\beta_0 = 0.9123828$ ,  $\beta_1 = -0.6522518$ ,  $\beta_2 = -0.9434133$ ,  $\beta_3 = -1.5792121$ ,  $X_1 = \text{Enhanced-Phase-1-wavelet-LLL-Skewness}$ ,  $X_2 = \text{Skewness-wavelet-LLL-GLCM-IDMN}$ ,  $X_3 = \text{Skewness-GLSZM-LGLZE}$ . Rad-TILs score was defined as the probability of high TILs predicted by radiomics model ( $p$ ), where higher Rad-TILs score indicated a higher predictive probability of high TILs based on the three representative radiomics features. The efficiency of the prediction model was assessed by the receiver operating characteristic (ROC), specificity, sensitivity and accuracy in the validation cohort.

### Comparison of enriched pathways between groups

The Rad-TILs score was calculated by the Rad-TILs model in the subcohort of patients whose RNA-seq data were available ( $n=116$ ). Patients were separated into high- and low-Rad-TILs groups by the median Rad-TILs score in the model training process. We conducted differentially expressed gene (DEG) selection (“limma” package in R) [32] and KEGG pathway analysis (“clusterProfiler” package in R) [33]. Furthermore, we conducted gene set enrichment analysis (GSEA) using the KEGG and Reactome databases (“clusterProfiler” package in R) [33] to compare enriched pathways between high- and low-Rad-TILs patients.

### Comparison of immune infiltration in the microenvironment between groups

A previously published reference matrix of gene sets characterizing different immune cell populations suitable for breast cancer [34] was adopted in the present study. Single sample gene set enrichment analysis (ssGSEA) was used to calculate the immune cell abundance score in every patient (“GSVA” package in R) [35]. Then, the Wilcoxon test was utilized to compare the difference between the high- and low-Rad-TILs groups.

### Comparison of immune-related molecules between groups

Cytokines, costimulators, coinhibitors and major histocompatibility complexes (MHCs) were compared between the two groups by transcriptomic analysis. Furthermore, two gene signatures characterizing T cell inflammation status [36] and T cell cytolytic activity [37] were adopted to infer the T cell status in the two groups of patients.

### Statistical analysis

Student’s t test and Wilcoxon’s 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. All the tests were two sided.  $P < 0.05$  was regarded as indicating significance, and  $0.05 < P < 0.1$  was regarded as marginally significance unless otherwise stated. The false discovery rate (FDR) correction was used in multiple hypothesis testing to decrease false positive rates. All statistical analyses were performed with R software (version 4.0.3, <http://www.R-project.org>).

## Results

### TILs-related radiomics feature selection and prediction model establishment

We retrospectively curated 139 TNBC samples with preoperative DEC-MRI and post-operative H&E pathological slides to establish a TILs evaluation cohort. The intention of the study was split into two parts: generation of TILs prediction radiomics model and illustration of biological basis of the radiomics model (Fig. 1). A percentage of sTILs  $\geq 20\%$  was defined as high TILs level (Fig. 2A, B). With all the extracted radiomics features, we used elastic net regression to select the features that most closely correlated with tumor-infiltrating T lymphocytes (TILs) in the training cohort. The following three radiomics features were finally selected: Enhanced-Phase-1-wavelet-LLL-Skewness (spatial

domain feature) that describes a first order imaging feature after applying wavelet filtering transformation of original first post-enhanced phase images, Skewness-wavelet-LLL-GLCM-IDMN (time domain feature) that depicts the variance pattern of a textural feature after wavelet filtering between each enhanced phase, and Skewness-GLSZM-LGLZE (time domain feature) that reflects the variance pattern of a textural feature between each enhanced phase (detailed description of radiomics features was presented in Additional file 1). The correlation between selected radiomics features and clinical characteristics was listed in Additional file 1: Table S1. In the training and validation cohorts, the three features presented a relatively lower value in tumors with high TILs (Fig. 2C). Then, the radiomic features were used as variables for logistic regression to build a prediction model. A predicted score reflecting the probability of high or low TILs (Radiomics TILs score, Rad-TILs score) for each patient was generated. We used the median of the Rad-TILs scores as a cutoff value to discriminate distinct Rad-TILs levels (Fig. 2D, E). The area under the receiver operating characteristic (ROC) curve (AUC) was 0.868 (95% CI 0.797–0.938) when predicting TILs in the training cohort, and the AUC was 0.790 (95% CI 0.638–0.943) in the validation cohort (Fig. 2F, G). In addition, the performance of the prediction model was tested for specificity (0.70), sensitivity (0.89) and accuracy (0.71) in the validation cohort.

#### Immune-related pathways enriched in high Rad-TILs score patients

Cohort of patient with available radiomics and RNA-seq data was then used to investigate the transcriptomic difference between two sets of patients with distinct Rad-TILs levels. First, differentially expressed genes (DEGs) related to immunity, such as CXCL11, CXCL13, and IDO1, were discovered between the two groups (Fig. 3A). KEGG analysis inferred that several pathways correlated with the immune response, such as antigen processing ( $p=0.02$ ) and presentation and PD-L1 expression and PD-1 checkpoint pathway in cancer ( $p=0.04$ ), were significantly upregulated in high Rad-TILs patients (Fig. 3B). From the most upregulated and downregulated pathways summarized by GSEA based on the KEGG and Reactome databases, we found that the upregulated pathways were mainly enriched in immune response and immune modulation. However, the downregulated genes were difficult to categorize (Fig. 3C, D). Moreover, natural killer cell-mediated cytotoxicity and T cell receptor

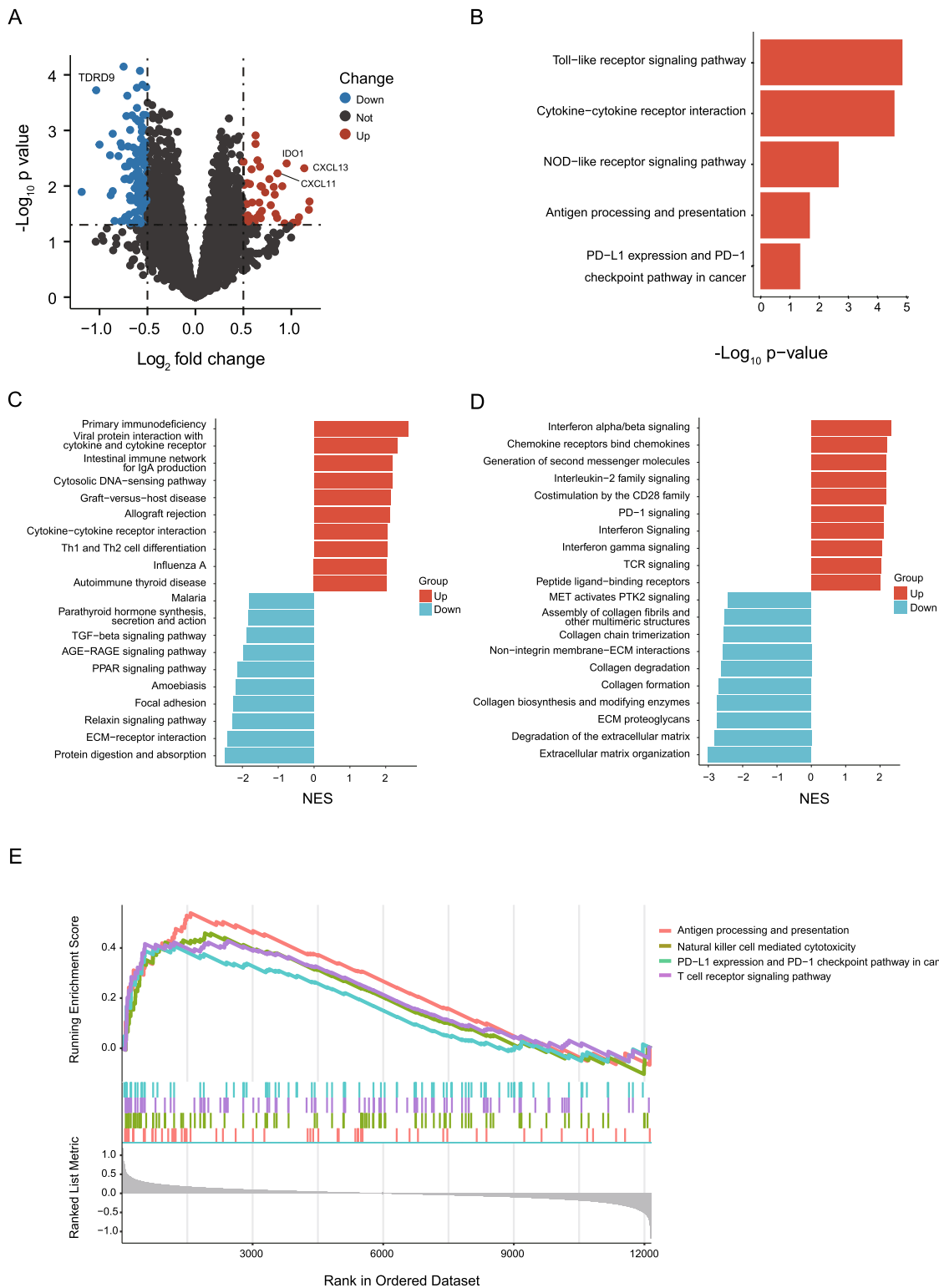
signaling pathway were upregulated in high Rad-TILs patients (Fig. 3E). Thus, we verified the different immune responses at the transcriptome level between the high- and low-Rad-TILs groups predicted by radiomics.

#### A hot immune microenvironment in high Rad-TILs score patients

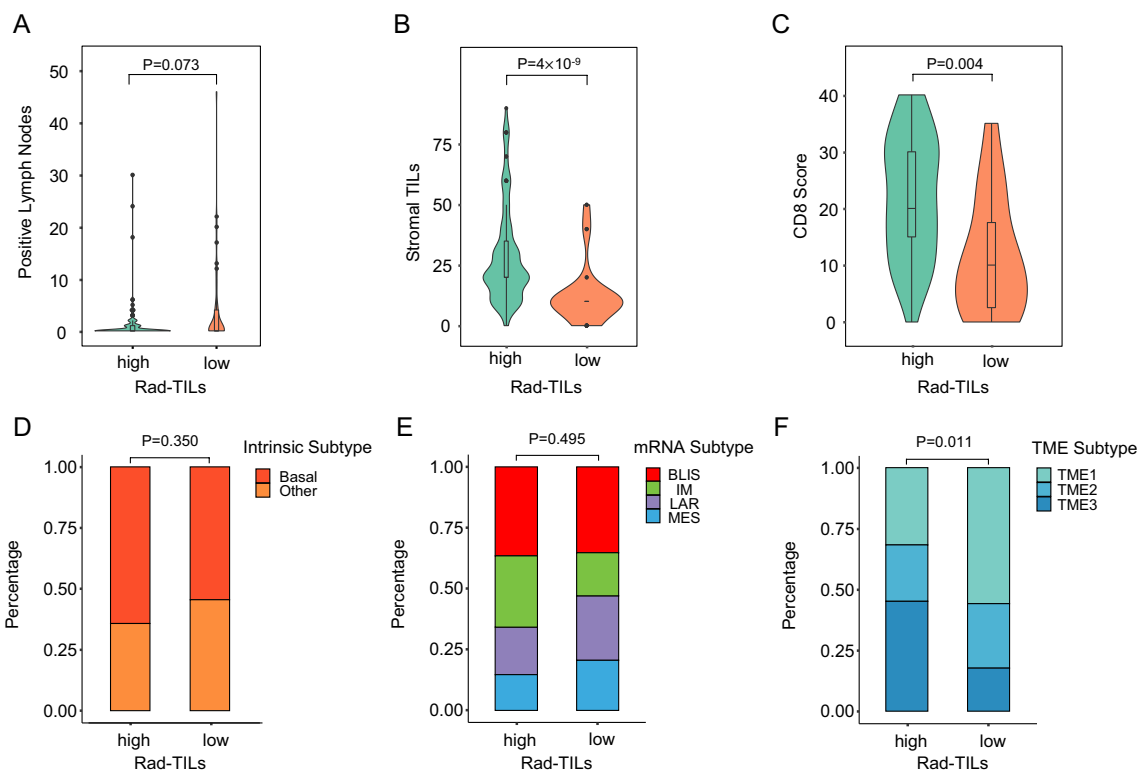
We analyzed the correlation between Rad-TILs score levels and the clinicopathological characteristics of TNBC patients. High Rad-TILs score patients tended to have fewer pathologically positive lymph nodes ( $p=0.073$ ), but the difference was not statistically significant (Fig. 4A). Higher stromal TILs (sTILs) and immunohistochemistry (IHC) CD8 scores were detected in high Rad-TILs score patients (Fig. 4B, C). The intrinsic subtypes, mRNA subtypes, fibrosis and necrosis were equivalent between the two groups (Fig. 4D, E). Furthermore, tumor microenvironment (TME) cluster 3, which was proposed in our previous study to characterize the inflammatory immune status of TNBC [34], was significantly enriched in high Rad-TILs score patients (Fig. 4F).

We also compared the difference in the composition of immune cells in the TME inferred by RNA-seq between high- and low-Rad-TILs score patients. Memory B cells, M1 macrophages, activated NK cells, plasma cells, CD8 T cells, follicular helper T cells and regulatory T cells were significantly or marginally significantly increased in the high Rad-TILs score group (Fig. 5A). Based on two published immune cell signatures, we found that patients with high Rad-TILs scores exhibited higher cytolytic activity and T cell inflamed gene expression profiles (Fig. 5B, C).

A relatively inflammatory TME in high Rad-TILs tumors was also indicated by the comparison of key molecules on the cell surface and cell-cell interactions. Several molecules expressed on the cell surface, including costimulators, coinhibitors and major histocompatibility complex (MHC), were highly expressed in high Rad-TILs score patient tumors (Fig. 5D, E, J). In addition, the levels of secreted immune-related cytokines, such as interleukins (ILs), colony-stimulating factors (CSFs), interferons (IFNs) and chemokines, were significantly elevated in the high group of patients (Fig. 5F–I), while the levels of transforming growth factors (TGFs) and tumor necrosis factors (TNFs) were equivalent between the two groups. Consequently, we established the relationship between opaque radiomics features and meaningful molecular features. Apart from distinct TILs levels, high Rad-TILs TNBC samples exhibited a hot immune microenvironment.



**Fig. 3** Immune related pathways enriched in the high TILs samples predicted by the radiomics model. **A** Differentially expressed genes between high and low TILs samples predicted by the radiomics model (high and low Rad-TILs). **B** Representative upregulated pathways in high Rad-TILs by KEGG enrichment analysis. **C** Top 10 upregulated and downregulated pathways in high Rad-TILs by GSEA based on KEGG database. **D** Top 10 upregulated and downregulated pathways in high Rad-TILs by GSEA based on Reactome databases. **E** Representative immune-related pathways enriched in high Rad-TILs by GSEA based on KEGG database



**Fig. 4** Clinicopathological and tumor microenvironmental (TME) characteristics of high and low Rad-TILs. **A, C** Lymph node status (**A**), sTILs quantification (**B**) and CD8 score (**C**) of high and low Rad-TILs. **D–F** Equivalent intrinsic subtypes (**D**), equivalent mRNA subtypes (**E**), and distinct TME subtypes (**F**) between high and low Rad-TILs. The p values were calculated using Wilcoxon test (for lymph node status, sTILs quantification and CD8 score) and Fisher's exact test (for intrinsic subtypes, mRNA subtypes and TME subtypes)

## Discussion

In the present study, we trained a TILs prediction model in the discovery cohort with a noninvasive radiomics method, which performed well in an additional validation cohort. In further investigation, we found a negative correlation between the Rad-TILs score and clinical risk factors, as well as the activated microenvironment exhibited in high Rad-TILs samples inferred by transcriptomics data, which supported and verified our initial finding. Significantly, radiomics combined with pathologic and transcriptomic data effectively reflected TILs status, and its potential mechanism was first reported in our study.

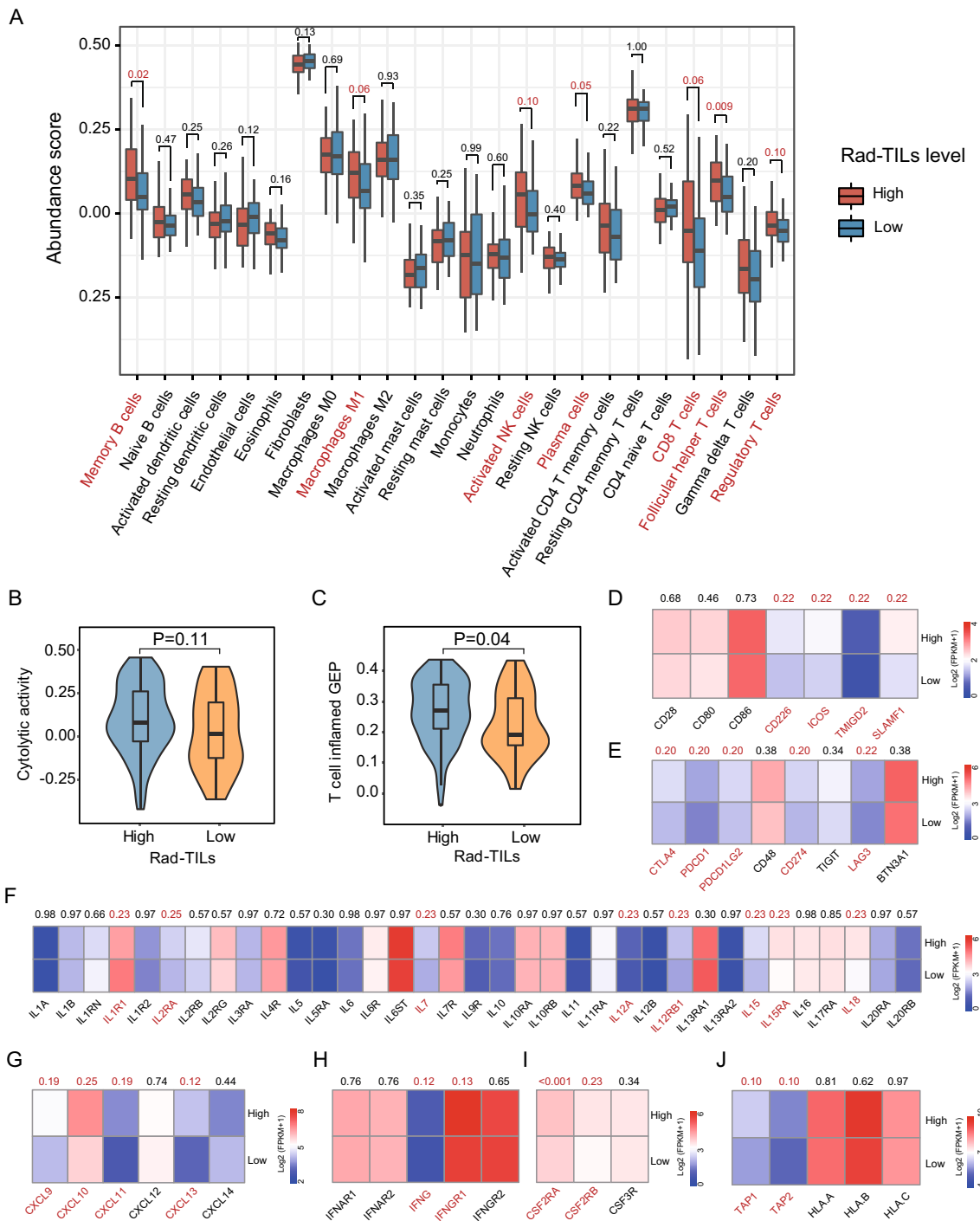
TILs play an important role in cancer biology and clinical oncology. TILs are regarded as biomarkers for immune infiltration and the prognosis of cancer patients and are promising potential biomarkers of patient response to immunotherapy. However, TILs quantification currently relies on manual evaluation of pathological slides, which is limited by the invasive method of specimen collection and time-consuming analysis approach.

Using a TNBC radiomics cohort with matched transcriptomic data, we established a three-feature radiomics signature, the Rad-TILs score, to noninvasively predict

the level of sTILs, which are more commonly measured clinically than intratumoral TILs (iTILs). The prediction model performed well in the validation cohort with an AUC of 0.79. In addition, the high accuracy (0.71), sensitivity (0.89) and specificity (0.70) also validated our model. Prior to the present study, several studies explored the relationship between radiomics and TILs [38–46]. Consistent with our work, the range of AUC, sensitivity and specificity of these studies were 0.67–0.87, 0.63–0.89 and 0.56–0.91, respectively. Interestingly, we found that most of the studies achieved high sensitivity but low specificity, which was also testified in our results. We thus speculate that radiomics in TILs prediction is a method with high sensitivity and low specificity.

Although previous studies have reported the predictive value of radiomics features in TILs prediction, the biological characteristics of these crucial radiomics features or image subgroups were not fully investigated. A deeper investigation of distinct image subgroups provided novel insight into the interpretability of opaque radiomics features and correlated molecular features apart from TILs. In this study, we revealed the distinct TME features of patients with high and low lymphocyte infiltration





**Fig. 5** Inflamed TME in high Rad-TILs revealed by transcriptomics analysis. **A** Comparison of immune cell subpopulation between high and low Rad-TILs. The p values were calculated using Wilcoxon test. Specific p values were denoted on each cell type. **B, C** Expression of immune-related signatures in high and low Rad-TILs. Cytolytic activity (**B**) and T cell inflamed gene expression profiles (GEPs) (**C**) of two groups. The p values were calculated using Wilcoxon test. **D, J** Distinct expression of immune-related molecules on cell surface, including costimulators (**D**), coinhibitors (**E**) and major histocompatibility complex (**J**) between two groups. Distinct immune-related secretory molecules, including interleukins (**F**), chemokines (**G**), interferons (**H**) and colony-stimulating factors (**I**) between two groups. The p values calculated by Wilcoxon test were adjusted to false discovery rate (FDR) using the Benjamini–Hochberg procedure in multiple comparisons. Specific FDR values were denoted on each molecule

predicted by radiomics (high and low Rad-TILs score) using matched RNA-seq, which demonstrated the unique value of multiomics in exploring the biological mechanism of radiomics features. First, we analyzed the DEGs between high- and low-Rad-TILs score patients and revealed that immune-related pathways were significantly enriched in the high-Rad-TILs score group. In addition, we compared the immune-related signatures, molecules and breast cancer subtypes between the two groups. Two representative signatures of the quantity and activity of T cells [36, 37, 47], several cytokines, immune checkpoint molecules and MHC molecules, were increased in the high Rad-TILs score group. The proportion of clusters characterizing immune inflammation [34] in the high Rad-TILs score group was also larger than that in the low Rad-TILs score group. Thus, it can be concluded that high Rad-TILs score tumors have an inflammatory immune microenvironment, and patients with high Rad-TILs scores are more likely to be sensitive to immunotherapy and have a better clinical outcome; however, the application of the Rad-TILs signature needs further validation in larger independent cohorts.

Moreover, we investigated the immune cells and TME subtypes in the two groups of patients. In addition to CD8-positive T cells, a variety of immune cells encompassing T helper cells, regulatory T cells, CD8 T cells, M1 macrophages, memory B cells and plasma cells aggregated in the high Rad-TILs score group. It has been reported that TILs comprise CD8<sup>+</sup> cytolytic T cells, CD4<sup>+</sup> helper T cells, CD20<sup>+</sup> B cells and NK cells [3, 48, 49], which is consistent with our prediction. The important role of T cells has been well established [50, 51], and recent studies have shed light on the function of B cells and plasma cells in the homeostasis of the TME [52]. Specifically, B cells promote antitumor immunity through antibody and cytokine production, antigen presentation and their role in tertiary lymphoid structures formation [52]. Kroeger et al. discovered that plasma cells were strongly associated with CD8<sup>+</sup> cytolytic T cells, and prognostic benefits were found only when coexisting with CD4<sup>+</sup>, CD20<sup>+</sup> TILs and plasma cells in ovarian cancer [53]. Consistent with the results of previous studies, our study revealed the important role of B cells in the TME.

Several limitations still remain in our research. First, the prediction model was built and tested based on a single-center radiomics cohort. The universality of the model remains to be externally validated. In addition, the transcriptomic analysis inferred distinct TMEs between high- and low-Rad-TILs samples. However, the results need further phenotypic characterization and mechanistic investigation.

We propose two future directions for further studies. First, multicenter and prospective clinical trial are necessary to demonstrate the generalization of TILs prediction model. Second, recent study revealed distinct TILs infiltration phenotypes in cancer termed immune inflamed, immune desert and immune excluded, which indicated that lymphocytes infiltration pattern but not the density of TILs determined the activating status of anti-tumor immunity [54]. Whether these immune phenotypes were more valuable than TILs density as a predictive biomarker needs to be explored in future studies.

In conclusion, we established a TILs prediction model using radiomics features in a TNBC radiomics cohort and revealed the distinct composition and characteristics of the microenvironment in two groups of patients differentiated by our radiomics model. The radiomics model is promising for application in clinical practice and may become a noninvasive biomarker for therapeutic stratification and prognostic prediction among TNBC patients.

#### Abbreviations

AUC: Area under the ROC curve; CSFs: Colony-stimulating factors; DEG: Differentially expressed gene; DFS: Disease-free survival; dMMR: Deficient mismatch repair; ER: Estrogen receptor; FDR: False discovery rate; FUSCC: Fudan University Shanghai Cancer Center; GSEA: Gene set enrichment analysis; H and E: Hematoxylin and eosin; HER2: Human epidermal growth factor receptor type 2; ICBs: Immune checkpoint blockades; IFNs: Interferons; IHC: Immunohistochemistry; ILs: Interleukins; iTILs: Intratumoral tumor-infiltrating lymphocytes; LPBC: Lymphocyte-predominant breast cancer; MHCs: Major histocompatibility complexes; MRI: Magnetic resonance imaging; MSI: Microsatellite instability; non-LPBC: Non-lymphocyte-predominant breast cancer; OS: Overall survival; PD-L1: Programmed cell death-ligand 1; PR: Progesterone receptor; RFS: Recurrence-free survival; RNA-seq: RNA sequencing; ROC: Receiver operating characteristic; ROIs: Regions of interest; ssGSEA: Single sample gene set enrichment analysis; sTILs: Stromal tumor-infiltrating lymphocytes; TGFs: Transforming growth factors; TILs: Tumor-infiltrating lymphocytes; TMB: Tumor mutation burden; TME: Tumor microenvironment; TNBC: Triple-negative breast cancer; TNFs: Tumor necrosis factors.

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-022-03688-x>.

**Additional file 1: Table S1.** Correlation between selected radiomics features and clinical characteristics listed in Table 1. **Table S2.** Feature categories used in this study. **Supplementary Methods.** Radiomics features calculation.

#### Acknowledgements

We thank the staff of the Radiology Department of Fudan University Shanghai Cancer Center for their assistance in breast MRI imaging collection. We thank Dr. Dan-Dan Zhang for her contribution to ROI delineation in this study. We thank Dr. Shen Zhao for his contribution to TILs evaluation and representative pathological slides selection. In addition, we thank the staff of the Institute of Science and Technology for Brain-inspired Intelligence of Fudan University for their contribution to radiomics feature extraction.

#### Author contributions

Conception and design: CY and Y-JG. Development of methodology: G-HS, YX, LJ and R-CZ. Acquisition of data: G-HS, LJ, and YC. Analysis and interpretation

of data: G-HS and YX. Writing, review, and/or revision of the manuscript: G-HS, YX, LJ, YC, CY, and Y-JG. Study supervision: Z-MS, CY and Y-JG. All authors read and approved the final manuscript.

### Funding

This project was supported by grants from The National Natural Science Foundation of China (81901703, 82071878, 82271957, 91959207 and 92159301), Shanghai Science and Technology Innovation Program (22Y11912700), Youth Medical Talents-Clinical Imaging Practitioner Program (SHWRS (2020) 087), and Clinical Research Plan of SHDC (SHDC2020CR2008A and SHDC12021103) and the SHDC Municipal Project for Developing Emerging and Frontier Technology in Shanghai Hospitals (SHDC12021103).

### Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the National Omics Data Encyclopedia (NODE), and can be viewed in NODE (<http://www.biosino.org/node>) by pasting the accession (OEP000155) into the text search box or through the URL:<http://www.biosino.org/node/project/detail/OEP000155>.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

Dr. Lin Jiang is currently an employee of AstraZeneca.

#### Author details

<sup>1</sup>Department of Breast Surgery, Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, No. 270 Dong'an Road, Shanghai 200032, People's Republic of China. <sup>2</sup>Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China. <sup>3</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 201203, China. <sup>4</sup>Division of Cancer and Stem Cell, School of Medicine at University of Nottingham, Nottingham, UK. <sup>5</sup>Department of Radiology, Fudan University Shanghai Cancer Center, No. 270 Dong'an Road, Shanghai 200032, People's Republic of China.

Received: 23 February 2022 Accepted: 6 October 2022

Published online: 15 October 2022

### References

- Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363:1938–48.
- Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*. 2016;13:674–90.
- Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, Loi S. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol*. 2016;13:228–41.
- Byrne A, Savas P, Sant S, Li R, Virassamy B, Luen SJ, Beavis PA, Mackay LK, Neeson PJ, Loi S. Tissue-resident memory T cells in breast cancer control and immunotherapy responses. *Nat Rev Clin Oncol*. 2020;17:341–8.
- Loi S, Sirtaine N, Piette F, Salgado R, Viale G, Van Eenoo F, Rouas G, Francis P, Crown JP, Hitre E, et al. Prognostic and predictive value of tumor-infiltrating lymphocytes in a phase III randomized adjuvant breast cancer trial in node-positive breast cancer comparing the addition of docetaxel to doxorubicin with doxorubicin-based chemotherapy: BIG 02–98. *J Clin Oncol*. 2013;31:860–7.
- Ali HR, Provenzano E, Dawson SJ, Blows FM, Liu B, Shah M, Earl HM, Poole CJ, Hiller L, Dunn JA, et al. Association between CD8+ T-cell infiltration and breast cancer survival in 12,439 patients. *Ann Oncol*. 2014;25:1536–43.
- Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol*. 2014;32:2959–66.
- Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C, et al. Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol*. 2014;25:1544–50.
- Pruneri G, Vingiani A, Bagnardi V, Rotmensz N, De Rose A, Palazzo A, Colleoni AM, Goldhirsch A, Viale G. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. *Ann Oncol*. 2016;27:249–56.
- Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, et al. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol*. 2010;28:105–13.
- Denkert C, von Minckwitz G, Brase JC, Sinn BV, Gade S, Kronenwett R, Pfitzner BM, Salat C, Loi S, Schmitt WD, et al. Tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy with or without carboplatin in human epidermal growth factor receptor 2-positive and triple-negative primary breast cancers. *J Clin Oncol*. 2015;33:983–91.
- Issa-Nummer Y, Darb-Esfahani S, Loibl S, Kunz G, Nekljudova V, Schrader I, Sinn BV, Ulmer HU, Kronenwett R, Just M, et al. Prospective validation of immunological infiltrate for prediction of response to neoadjuvant chemotherapy in HER2-negative breast cancer—a substudy of the neoadjuvant GeparQuinto trial. *PLoS ONE*. 2013;8:e79775.
- Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, Denkert C, Park YH, Hui R, Harbeck N, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med*. 2020;382:810–21.
- Schmid P, Rugo HS, Adams S, Schneeweiss A, Barrios CH, Iwata H, Diéras V, Henschel V, Molinero L, Chui SY, et al. Atezolizumab plus nab-paclitaxel as first-line treatment for unresectable, locally advanced or metastatic triple-negative breast cancer (IMpassion130): updated efficacy results from a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2020;21:44–59.
- Mittendorf EA, Zhang H, Barrios CH, Saji S, Jung KH, Hegg R, Koehler A, Sohn J, Iwata H, Telli ML, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet*. 2020;396:1090–100.
- Goodman AM, Sokol ES, Frampton GM, Lippman SM, Kurzrock R. Microsatellite-stable tumors with high mutational burden benefit from immunotherapy. *Cancer Immunol Res*. 2019;7:1570–3.
- Marabelle A, Le DT, Ascierto PA, Di Giacomo AM, De Jesus-Acosta A, Delord JP, Geva R, Gottfried M, Penel N, Hansen AR, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J Clin Oncol*. 2020;38:1–10.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures they are data. *Radiology*. 2016;278:563–77.
- Grossmann P, Stringfield O, El-Hachem N, Bui MM, Rios Velazquez E, Parmar C, Leijenaar RT, Haibe-Kains B, Lambin P, Gillies RJ, Aerts HJ. Defining the biological basis of radiomic phenotypes in lung cancer. *Elife*. 2017;6:e23421.
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
- Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, Ma ZL, Liu ZY. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34:2157–64.
- Kong J, Zheng J, Wu J, Wu S, Cai J, Diao X, Xie W, Chen X, Yu H, Huang L, et al. Development of a radiomics model to diagnose

- pheochromocytoma preoperatively: a multicenter study with prospective validation. *J Transl Med.* 2022;20:31.
24. Sun K, Jiao Z, Zhu H, Chai W, Yan X, Fu C, Cheng JZ, Yan F, Shen D. Radiomics-based machine learning analysis and characterization of breast lesions with multiparametric diffusion-weighted MR. *J Transl Med.* 2021;19:443.
  25. Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, Wienert S, Van den Eynden G, Baehner FL, Penault-Llorca F, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol.* 2015;26:259–71.
  26. Hasebe T, Tsuda H, Hirohashi S, Shimosato Y, Tsubono Y, Yamamoto H, Mukai K. Fibrotic focus in infiltrating ductal carcinoma of the breast: a significant histopathological prognostic parameter for predicting the long-term survival of the patients. *Breast Cancer Res Treat.* 1998;49:195–208.
  27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159–74.
  28. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008;12:26–41.
  29. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging.* 1998;17:87–97.
  30. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, Beets-Tan RGH, Fillion-Robin JC, Pieper S, Aerts H. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77:e104–7.
  31. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc: Ser B.* 2005;67:301–20.
  32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43:e47.
  33. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16:284–7.
  34. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, Ruan M, Wang H, Zhao J, Li Q, et al. Multi-omics profiling reveals distinct microenvironment characterization and suggests immune escape mechanisms of triple-negative breast cancer. *Clin Cancer Res.* 2019;25:5002–14.
  35. Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013;14:7.
  36. Ott PA, Bang YJ, Piha-Paul SA, Razak ARA, Bennouna J, Sorria JC, Rugo HS, Cohen RB, O'Neil BH, Mehnert JM, et al. T-cell-inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J Clin Oncol.* 2019;37:318–27.
  37. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell.* 2015;160:48–61.
  38. Bian Y, Liu C, Li Q, Meng Y, Liu F, Zhang H, Fang X, Li J, Yu J, Feng X, et al. Preoperative radiomics approach to evaluating tumor-infiltrating CD8(+) T cells in patients with pancreatic ductal adenocarcinoma using noncontrast magnetic resonance imaging. *J Magn Reson Imaging.* 2022;55:803–14.
  39. Bian Y, Liu YF, Jiang H, Meng Y, Liu F, Cao K, Zhang H, Fang X, Li J, Yu J, et al. Machine learning for MRI radiomics: a study predicting tumor-infiltrating lymphocytes in patients with pancreatic ductal adenocarcinoma. *Abdom Radiol.* 2021;46:4800–16.
  40. Li J, Shi Z, Liu F, Fang X, Cao K, Meng Y, Zhang H, Yu J, Feng X, Li Q, et al. XGBoost classifier based on computed tomography radiomics for prediction of tumor-infiltrating CD8(+) T-cells in patients with pancreatic ductal adenocarcinoma. *Front Oncol.* 2021;11:671333.
  41. Liao H, Zhang Z, Chen J, Liao M, Xu L, Wu Z, Yuan K, Song B, Zeng Y. Preoperative radiomic approach to evaluate tumor-infiltrating CD8(+) T cells in hepatocellular carcinoma patients using contrast-enhanced computed tomography. *Ann Surg Oncol.* 2019;26:4537–47.
  42. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, Verlingue L, Brandao D, Lancia A, Ammari S, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* 2018;19:1180–91.
  43. Tang WJ, Kong QC, Cheng ZX, Liang YS, Jin Z, Chen LX, Hu WK, Liang YY, Wei XH, Guo Y, Jiang XQ. Performance of radiomics models for tumour-infiltrating lymphocyte (TIL) prediction in breast cancer: the role of the dynamic contrast-enhanced (DCE) MRI phase. *Eur Radiol.* 2022;32:864–75.
  44. Xu N, Zhou J, He X, Ye S, Miao H, Liu H, Chen Z, Zhao Y, Pan Z, Wang M. Radiomics model for evaluating the level of tumor-infiltrating lymphocytes in breast cancer based on dynamic contrast-enhanced MRI. *Clin Breast Cancer.* 2021;21:440-449.e441.
  45. Yu H, Meng X, Chen H, Han X, Fan J, Gao W, Du L, Chen Y, Wang Y, Liu X, et al. Correlation between mammographic radiomics features and the level of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer. *Front Oncol.* 2020;10:412.
  46. Yu H, Meng X, Chen H, Liu J, Gao W, Du L, Chen Y, Wang Y, Liu X, Liu B, et al. Predicting the level of tumor-infiltrating lymphocytes in patients with breast cancer: usefulness of mammographic radiomics features. *Front Oncol.* 2021;11:628577.
  47. Ayers M, Luceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V, et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest.* 2017;127:2930–40.
  48. Whitford P, George WD, Campbell AM. Flow cytometric analysis of tumour infiltrating lymphocyte activation and tumour cell MHC class I and II expression in breast cancer patients. *Cancer Lett.* 1992;61:157–64.
  49. Chin Y, Janseens J, Vandepitte J, Vandenbrande J, Opdebeek L, Raus J. Phenotypic analysis of tumor-infiltrating lymphocytes from human breast cancer. *Anticancer Res.* 1992;12:1463–6.
  50. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol.* 2015;15:486–99.
  51. Borst J, Ahrends T, Babala N, Melief CJM, Kastenmuller W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* 2018;18:635–47.
  52. Sharonov GV, Serebrovskaya EO, Yuzhakova DV, Britanova OV, Chudakov DM. B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat Rev Immunol.* 2020;20:294–307.
  53. Kroeger DR, Milne K, Nelson BH. Tumor-infiltrating plasma cells are associated with tertiary lymphoid structures, cytolytic T-cell responses, and superior prognosis in ovarian cancer. *Clin Cancer Res.* 2016;22:3005–15.
  54. Park S, Ock CY, Kim H, Pereira S, Park S, Ma M, Choi S, Kim S, Shin S, Aum BJ, et al. Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. *J Clin Oncol.* 2022;40:1916–28.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

