

RESEARCH

Open Access



# Radiogenomic analysis of cellular tumor-stroma heterogeneity as a prognostic predictor in breast cancer

Ming Fan<sup>1</sup>, Kailang Wang<sup>1</sup>, You Zhang<sup>1</sup>, Yuanyuan Ge<sup>1</sup>, Zhong Lü<sup>2\*</sup> and Lihua Li<sup>1\*</sup> 

## Abstract

**Background** The tumor microenvironment and intercellular communication between solid tumors and the surrounding stroma play crucial roles in cancer initiation, progression, and prognosis. Radiomics provides clinically relevant information from radiological images; however, its biological implications in uncovering tumor pathophysiology driven by cellular heterogeneity between the tumor and stroma are largely unknown. We aimed to identify radiogenomic signatures of cellular tumor-stroma heterogeneity (TSH) to improve breast cancer management and prognosis analysis.

**Methods** This retrospective multicohort study included five datasets. Cell subpopulations were estimated using bulk gene expression data, and the relative difference in cell subpopulations between the tumor and stroma was used as a biomarker to categorize patients into good- and poor-survival groups. A radiogenomic signature-based model utilizing dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) was developed to target TSH, and its clinical significance in relation to survival outcomes was independently validated.

**Results** The final cohorts of 1330 women were included for cellular TSH biomarker identification ( $n = 112$ , mean age, 57.3 years  $\pm$  14.6) and validation ( $n = 886$ , mean age, 58.9 years  $\pm$  13.1), radiogenomic signature of TSH identification ( $n = 91$ , mean age, 55.5 years  $\pm$  11.4), and prognostic ( $n = 241$ ) assessments. The cytotoxic lymphocyte biomarker differentiated patients into good- and poor-survival groups ( $p < 0.0001$ ) and was independently validated ( $p = 0.014$ ). The good survival group exhibited denser cell interconnections. The radiogenomic signature of TSH was identified and showed a positive association with overall survival ( $p = 0.038$ ) and recurrence-free survival ( $p = 3 \times 10^{-4}$ ).

**Conclusion** Radiogenomic signatures provide insights into prognostic factors that reflect the imbalanced tumor-stroma environment, thereby presenting breast cancer-specific biological implications and prognostic significance.

**Keywords** Breast cancer, Cell subpopulation, Prognosis, Radiogenomics

## Introduction

Breast cancer exhibits genetic diversity both between and within tumors, leading to diverse tumor phenotypes, disease progression, and therapeutic resistance [1, 2]. The tumor microenvironment is a complex ecosystem comprising immune cells, fibroblasts, extracellular matrix, and cytokines [3], while the surrounding stromal cells [4, 5] protect tumor cells from immune activity [6]. Altered environmental heterogeneity influences tumor initiation, progression, and response to therapy [7]. Increased

\*Correspondence:

Zhong Lü

lz3399@wmu.edu.cn

Lihua Li

lih@hdu.edu.cn

<sup>1</sup> Institute of Intelligent Biomedicine, Hangzhou Dianzi University, Hangzhou 310018, China

<sup>2</sup> Affiliated Dongyang Hospital of Wenzhou Medical University, Dongyang 322100, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

tumor-infiltrating lymphocytes (TILs) have been found to be associated with chemotherapy response and improved survival [8]. Environmental heterogeneity, particularly in terms of immune cell subtypes such as CD4 cells and T cells, is associated with less activation of the immune response and poorer survival in breast cancer [9, 10]. A comprehensive understanding of tumor environmental heterogeneity is crucial for accurate prognostic assessment and effective treatment management of breast cancer.

To this end, many studies have focused on estimating the composition of multiple cell subpopulations by deconvoluting bulk tissue microarray data derived from tissue mixtures [11]. Technologies such as CIBERSORT have been developed to deconvolute large-scale RNA mixtures for the identification of cellular biomarkers and therapeutic targets [12]. By estimating the abundance of immune and stromal cell populations within heterogeneous tissues, environmental heterogeneity can be characterized [13]. Estimated cell subpopulations have been correlated with therapeutic response and survival outcomes in breast cancer [14–16]. Beyond this routine genomic approach, imaging techniques offer a noninvasive means to reveal tumor heterogeneity driven by underlying pathophysiological processes [17].

Currently, radiogenomic/radiomics studies are increasingly being performed by linking imaging features to molecular/genomic status [18–21] with the ultimate goal of improving disease management for patients [22–25]. Radiogenomic signatures of TILs have been identified and are associated with survival in patients with breast cancer [26]. Predictive imaging features that reflect cell subpopulations estimated from genomic data were identified to stratify the survival of patients [27, 28]. Tumor microenvironment-associated imaging features were used to assess chemotherapy responses and/or survival outcomes [29]. Breast cancer heterogeneity was evaluated by texture analysis of dynamic-contrast enhanced magnetic resonance imaging (DCE-MRI) and T2-weighted MRI for the prediction of survival outcomes in breast cancer [30]. All these studies demonstrated that tumor microenvironmental characteristics can be captured using noninvasive images that are ubiquitously used in clinical practice.

Emerging evidence suggests that the microenvironment is regulated through ongoing cellular crosstalk [31, 32], and interactions between tumor cells and the stroma play a crucial role in inducing an altered environment associated with cancer initiation, progression [2, 33], metastasis [34], response to neoadjuvant chemotherapy (NACT), and patient prognosis [35, 36]. While existing radiogenomic studies have predominantly focused on analyzing either tumors or the surrounding stroma, the

predictive value of imaging in capturing the cellular heterogeneity between the two, which serves as a surrogate for prognosis analysis, remains uncertain.

The purpose of our study was to identify radiogenomic signatures of cellular tumor-stroma heterogeneity (TSH) and to investigate the potential of imaging TSH as a biomarker for predicting prognosis in breast cancer. In contrast to conventional data-driven studies that focus on identifying clinically relevant imaging signatures, our radiogenomic analysis delves into the cellular interactions between tumor and stroma, thereby providing additional genomic-level information with biological significance.

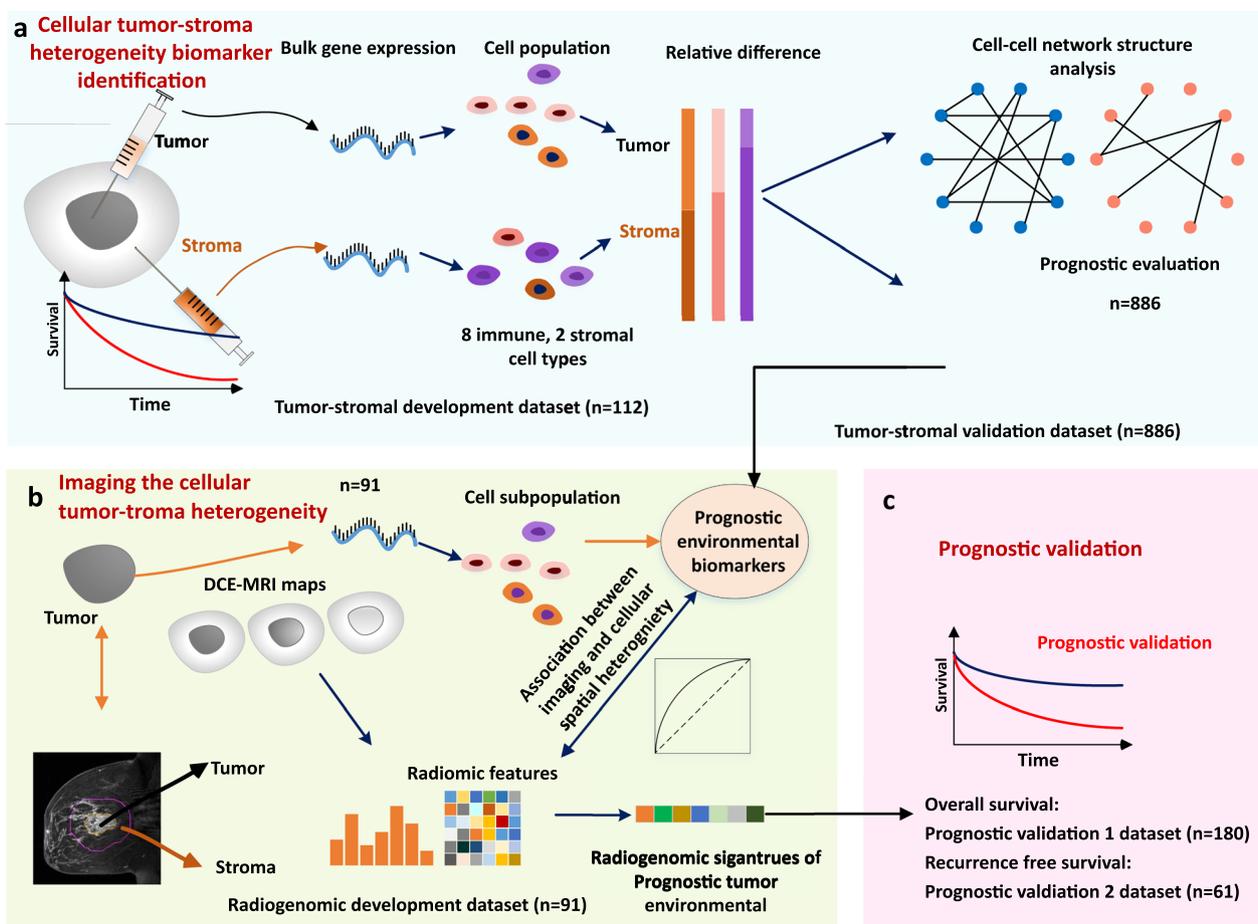
## Materials and methods

### Study design

The framework of this three-stage study is depicted in Fig. 1. First, we estimated cell subpopulations using the microenvironment cell population counter (MCP)-counter algorithm [13]. This was accomplished by utilizing bulk gene expression data from both the tumor and surrounding stroma, forming the tumor-stroma development dataset (Fig. 1a). The cellular TSH biomarker was defined as the relative difference in cell subpopulation abundance between the tumor and stroma. Patients were separated into good- and poor-survival groups according to their cellular TSH biomarker values. Moreover, we conducted an analysis of network-level cellular connections and correlations, wherein each cell subpopulation was represented as a node, and the relationships between these nodes were defined as edges.

It is important to note that to identify the radiogenomic signature of cellular TSH, it is necessary to have matched pairs of imaging and genomic data derived from both the tumor and stroma in the tumor-stroma development dataset. Unfortunately, such paired data are relatively scarce compared to data obtained solely from tumors, which presents a challenge in conducting radiogenomic analysis of TSH. Therefore, we mapped the predicted TSH using cell subpopulations derived exclusively from the tumor (Fig. 2). To validate the feasibility of our approach, we employed the tumor-stromal validation dataset comprising genomic data and follow-up data. The prognostic significance of TSH was confirmed in this dataset.

Second, the cellular biomarker-based predictive model, established using the tumor-stroma development dataset, was applied to the radiogenomic development dataset, which comprised matched imaging and genomic data (Fig. 1b). The TSH score was predicted by utilizing the estimated cell subpopulation data from the tumors. We extracted radiomic features ( $n=572$ ) from each sample and identified radiogenomic signatures by linking the imaging features derived from both the tumor



**Fig. 1** Study framework. **a** Cellular tumor-stroma heterogeneity biomarker identification. **b** Imaging the cellular tumor-stroma heterogeneity. **c** Prognostic validation. Radiogenomic analysis by associating cellular tumor-stroma biomarkers and imaging features

and surrounding stroma with the predicted cellular TSH values. An imaging signature-based model was subsequently developed to classify patients into either good or poor survival groups.

Third, we independently validated the prognostic significance of the identified radiogenomic signatures by assessing their association with survival in prognostic validation datasets 1 and 2 (Fig. 1c).

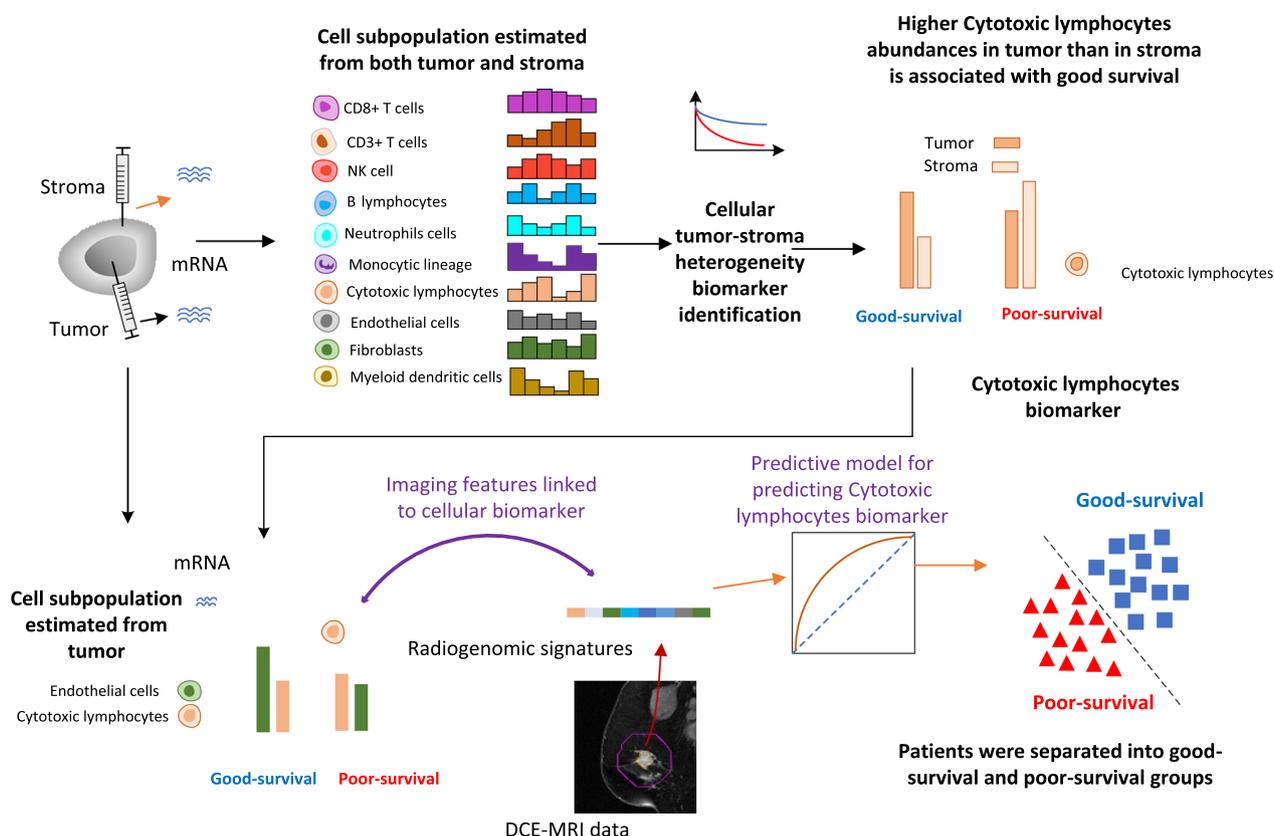
### Datasets

This retrospective study included multicohort datasets of 1330 patients with genomic, imaging, and clinical data (Fig. 3). The tumor-stromal development dataset (n=112) was collected from TCGA-BRCA [37, 38] of the Cancer Imaging Archive (TCIA) cohort. Initially, the dataset consisted of 1092 gene expression data points. We selected samples that had gene expression data from both the tumor and the surrounding parenchyma to form the tumor-stromal development dataset. This dataset was utilized to identify prognostic cell subpopulation

biomarkers associated with tumor-stroma heterogeneity (TSH) and to assess cell subpopulation connectivity within the tumor or stroma.

For the tumor-stromal validation dataset (n=886), we collected genomic data from tumors alone from TCGA-BRCA. The dataset initially included 1092 samples, while the samples overlapped in the tumor-stromal development dataset (n=112) and the radiogenomic development dataset (n=91) were eliminated. After excluding one patient with gene expression data solely from the parenchyma region, one patient without clinical data, and one with no available survival data, we obtained 886 samples with RNA data from tumors. This dataset served as a validation set to confirm the identified cell subpopulation biomarkers from the tumor-stroma development dataset.

The radiogenomic development dataset (n=91) was obtained from TCGA-BRCA with matched imaging and genomic data. This dataset initially included 137 samples. After excluding one sample with unavailable genomic



**Fig. 2** Imaging cellular tumor-stroma heterogeneity

data and 21 samples with incomplete imaging data, we retained 115 women for analysis. We excluded 24 samples with genomic data from both the tumor and surrounding stroma, as they were utilized in the tumor-stromal development dataset. The resulting dataset (n=91) contained tumor gene expression, matched imaging, and survival outcome data from the same patients. This dataset enabled the identification of radiogenomic signatures by linking imaging features with TSH scores.

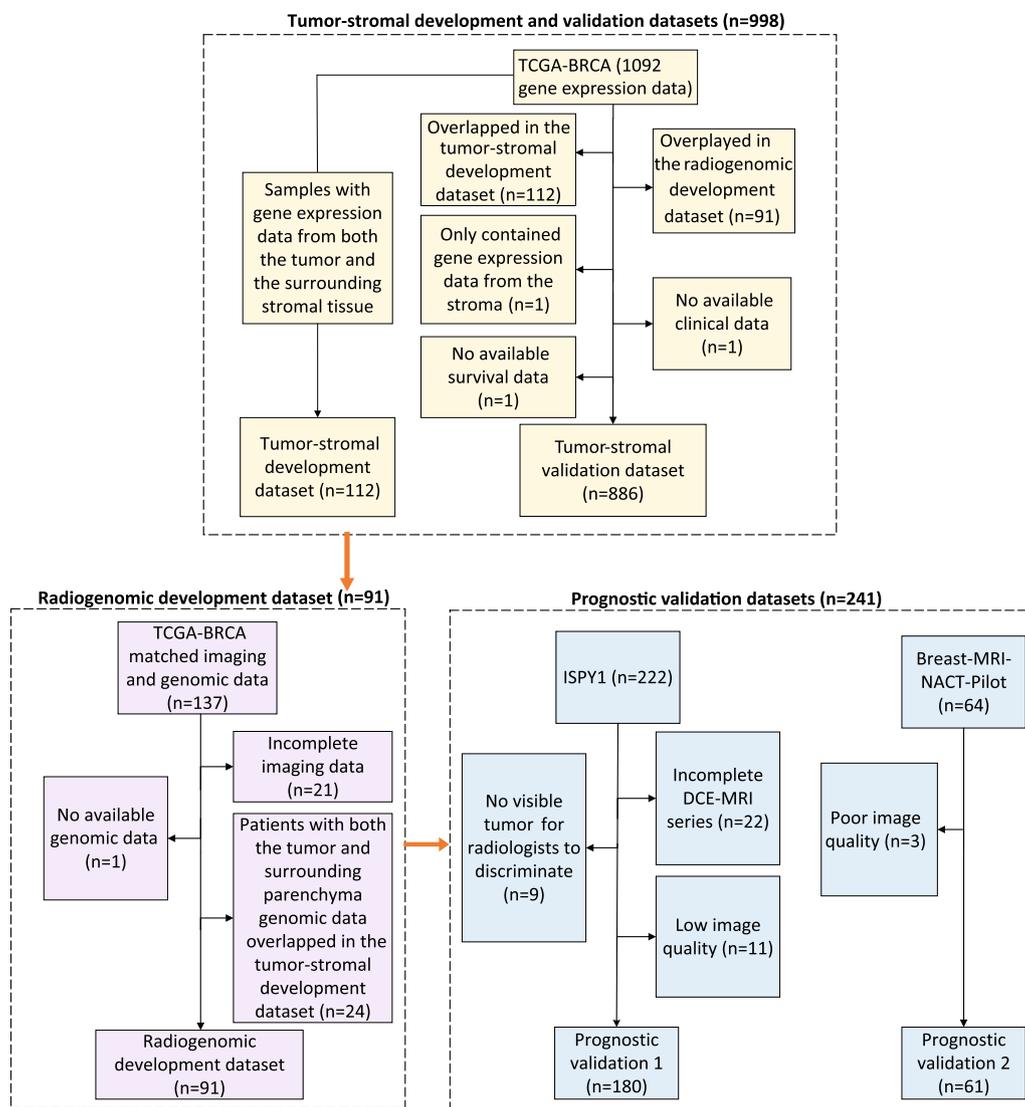
The prognostic significance of the radiogenomic signatures was validated using two datasets with imaging data and corresponding survival information. The prognostic validation 1 (n=61) and prognostic validation 2 (n=180) datasets were collected from the Breast-MRI-NACT-Pilot [39] and ISPY1 [40] datasets, respectively. The prognostic validation 1 dataset initially included 64 patients. After excluding three patients owing to poor image quality, we included 61 samples for analysis. The prognostic validation 2 dataset, initially consisting of 222 women, underwent additional exclusions: 22 patients were excluded due to incomplete DCE-MRI series, nine patients were excluded because the tumors were not visually distinguishable by radiologists, and 11 patients were

excluded due to poor image quality. The remaining dataset included 180 patients.

**Imaging protocols**

The radiogenomic development dataset contained DCE-MRI with a T1-weighted, 3D spoiled gradient echo sequence. The imaging protocol details can be found elsewhere [20]. The DCE-MRI comprised a precontrast image series (S<sub>0</sub>) and three to six postcontrast image series, with a temporal resolution of approximately 110 s. The in-plane resolution ranged from 0.53–0.86 mm, the matrix was 256×192, the slice thickness ranged from 2.0–3.0 mm, and the flip angle was 10°.

For the prognostic validation 1 dataset, DCE-MRI was conducted using a 1.5T fat-suppressed MRI scanner (GE Healthcare, Milwaukee, Wisconsin, USA), with patients positioned in the prone position. Further information regarding the imaging protocols can be found elsewhere [41]. The imaging parameters were as follows: repetition time (TR)=8 ms; echo time (TE)=8 ms; flip angle=20°; in-plane resolution ranging from 0.7–0.9 mm; and slice thickness ranging from 2–2.4 mm. After the administration of a bolus of 0.1 mmol/kg gadobutrol, the first and



**Fig. 3** Data collection flowchart

second postcontrast image series were acquired at 2.5 and 7.5 min, respectively.

For the prognostic validation 2 dataset, DCE-MRI was performed using a 1.5T scanner with fat-suppressed T1-weighted imaging. The imaging parameters were set as follows: matrix=256×192; field of view=16–18 cm, TR≤20 ms; TE=4.5 ms, flip angle=45°, slice thickness≤3 mm, and in-plane resolution≤1 mm. The post-contrast image series was acquired at 2.5 and 7.5 min after the administration of the contrast agent.

**Image preprocessing and feature analysis**

Owing to the varied imaging protocols of the different manufacturers used in the study, each image was resized

to a resolution of 0.8×0.8 mm and a thickness of 2 mm to ensure consistent image resolution. Subsequently, image normalization was performed by mapping each image to a standardized grayscale and dividing the pixel values by the mean value of the parenchymal region. Breast areas were obtained by excluding skin and chest walls. The stroma area was manually annotated and defined as a tissue band with a width of 20 mm (25×0.8 mm=20 mm) extending from the tumor boundary [42]. The segmentation of breast tumors was performed using a spatial fuzzy c-means clustering method with the tumor’s center location as the initial “seed” being manually annotated by two experienced radiologists in consensus, each possessing 10 years of expertise [43]. To ensure the accuracy of the

breast tumor segmentation results, manual correction was performed by the radiologists, which accounted for less than 5% of the segmentations.

Radiomic features (n=107) were calculated using the publicly available Pyradiomics software package [44]. These features comprised statistical (n=18), morphological (n=14) and textural features (n=75). The texture features included a gray-level cooccurrence matrix (GLCM) (n=24), gray-level run length matrix (GLRLM, n=16), gray-level zone length matrix (GLSZM, n=16), gray-level difference matrix (GLDM, n=14), and neighbourhood gray-tone difference matrix (NGTDM, n=5). A detailed list of these features is presented in Additional file 1: Table S1.

Imaging features were calculated from the tumor and surrounding stromal tissues at different imaging phases and maps, including precontrast ( $S_0$ ), the subtraction image between the intermediate postcontrast image and  $S_0$  (termed  $S_I$ ), and subtraction images between the last postcontrast image and  $S_0$  (termed  $S_L$ ). Morphological features, on the other hand, were only calculated on a single sequence from the tumors since there were no discernible shape characteristics present in the stromal regions. In total, 572 radiogenomic imaging features were extracted from each sample.

#### Cell subpopulation estimation and cell network analysis

The microenvironment cell population counter (MCP-counter) R package (<http://github.com/ebecht/MCPcounter>) was used to generate the absolute abundance scores of the immune and stromal cell types. The MCP-counter algorithm enables the estimation of eight immune cell types, namely, CD3+ T cells, CD8+ T cells, cytotoxic lymphocytes, NK cells, B lymphocytes, monocytic lineage, myeloid dendritic cells, and neutrophils, as well as two stromal cells, endothelial cells and fibroblasts, for comprehensive characterization of the tumor microenvironment [13]. This method was performed by comparing the gene expression profiles with predefined reference gene sets.

To ensure robust analysis, we excluded genes that displayed zero expression or were missing data in 80% of the samples. The estimation of each cell subpopulation was performed using log<sub>2</sub>-transformed gene expression data obtained from tumor and/or surrounding stromal samples.

In addition to investigating tumor-intrinsic or stromal factors, we also explored the influence of cellular biomarkers reflecting TSH, which may induce significant phenotypic alterations and influence the progression of breast cancer. The cell subpopulation biomarker associated with TSH was defined as the ratio of cell

subpopulation abundance between the tumor and stromal tissue, calculated using the following formula:

$$\frac{A_{Tumor} - A_{Stroma}}{A_{Stroma}},$$

where  $A_{Tumor}$  and  $A_{Stroma}$  denote the abundance of the cell subpopulations.

In the analysis of cellular networks, a cell subpopulation was designated as a node, while the establishment of network edges was based on node similarities under a predetermined threshold (Pearson correlation coefficient > 0.5). The evaluation of network topology in the good- and poor-survival groups involved the computation of various topological parameters, including network diameter, edge number, node number, network centralization, network heterogeneity, network density, clustering coefficient, and characteristic path length. These topological parameters were calculated using widely accessible Cytoscape software (<https://cytoscape.org/>) to ensure accurate measurements and consistent analysis methodologies.

#### Statistical analysis

Radiogenomic signatures were established by associating imaging data with predicted TSH scores using a random forest-based classifier. To mitigate multicollinearity effects, feature pairs with high similarity ( $r > 0.8$ ) were removed from the predictive model. Genetic algorithm-based feature selection was employed to generate an optimal subset of features for classification. A tenfold cross-validation procedure was used to optimize feature selection and model building, reducing the risk of underfitting or overfitting.

Survival analysis was performed using the Kaplan–Meier method, obtaining hazard ratios (HRs) with 95% confidence intervals. The log-rank test was utilized to evaluate differences in survival rates between stratified groups and determine the significance of the survival curves. A multivariate Cox regression model was employed to assess the independent association of the cell subpopulations or radiogenomic signatures with overall survival (OS) or recurrence-free survival (RFS) while adjusting for the available clinical variables, including age, estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and tumor size. Patients were censored at 10 years in the absence of events. A likelihood ratio test was performed to determine whether the inclusion of radiogenomic signatures as explanatory variables significantly improved model fit compared to models based solely on clinical variables.

The Benjamin–Hochberg method was used to control for false discovery. The Wilcoxon signed-rank test

was used to assess the differences in cell subpopulation abundance between the tumor and stroma. All statistical analyses were performed using MATLAB (R2018, MathWorks) and R (version 4.0; R Foundation for Statistical Computing).

## Results

### Study cohort

Five datasets comprising genomic, imaging, and clinical data were utilized for the identification and validation of TSH biomarkers, development of radiogenomic signatures, and prognostic assessments. Detailed descriptions of patient characteristics according to age, menopausal status, race, ER status, PR status, HER2 status, family history, tumor size, histological status, recurrence/survival status, and follow-up status (alive, dead, or lost to follow-up) are listed in Table 1.

### Cellular tumor-stroma heterogeneity biomarker identification

The estimated cell subpopulations from tumors, including T cells, CD8 T cells, cytotoxic lymphocytes, B lineage cells, and myeloid dendritic cells, exhibited a significant positive association with good OS (corrected  $p < 0.05$ ) (Additional file 1: Table S2). Conversely, the neutrophil cell subpopulation from the surrounding nontumoral tissue was negatively correlated with poor OS (corrected  $p < 0.05$ ) (Additional file 1: Table S3), consistent with previous research [45].

Univariate analysis of the cellular TSH biomarker revealed significant positive associations between T cells, cytotoxic lymphocytes, B-lineage cells, NK cells, myeloid dendritic cells, and good OS (corrected  $p < 0.05$ ). A higher abundance of these cell subpopulations within the tumor, compared to the stroma, was indicative of favorable survival outcomes (Table 2). Notably, the cytotoxic lymphocyte biomarker retained its independent association with breast cancer patient survival in the multivariable

**Table 1** Patient characteristics

Parameter	Tumor-stromal development (n = 112)	Tumor-stromal validation (n = 886)	Radiogenomic development (n = 91)	Prognostic validation 1 (n = 180)	Prognostic validation 2 (n = 61)
Age					
Median	56.5 (30–90)	59 (26–90)	56 (29–82)	42.9 (26.7–68.8)	48 (29.7–72.4)
Mean ± SD	57.3 ± 14.6	58.9 ± 13.1	55.5 ± 11.4	47.7 ± 8.8	48.1 ± 9.8
ER					
Positive	78 (70)	650 (73)	75 (82)	101 (56)	28 (46)
Negative	21 (19)	200 (23)	16 (18)	77 (43)	20 (33)
N/A	13 (11)	36 (4)	0	2 (1)	13 (21)
PR					
Positive	68 (61)	560 (63)	66 (73)	84 (47)	22 (36)
Negative	31 (28)	287 (32)	25 (27)	94 (52)	26 (43)
N/A	13 (11)	39 (5)	0	2 (1)	13 (21)
HER2					
Positive	23 (21)	128 (14)	13 (14)	52 (29)	14 (23)
Negative	56 (50)	459 (52)	44 (48)	124 (69)	31 (51)
N/A	33 (29)	299 (34)	34 (38)	4 (2)	16 (26)
Follow-up (years)					
Median	3.45 (0.0–10.8)	1.98 (0.0–23.6)	2.99 (0.5–9.4)	3.90(0.5–6.8)	5.39 (0.3–9.8)
Mean ± SD	3.92 ± 2.37	3.23 ± 3.39	3.66 ± 2.11	3.84 ± 1.45	4.77 ± 2.75
Recurrence					
Event	NA	NA	NA	49(27)	23(38)
No-event	NA	NA	NA	131(73)	38(62)
Death					
Event	43 (38)	97 (11)	2 (2)	33 (18)	NA
No-event	69 (62)	789 (89)	89 (98)	143 (80)	NA
Unknown	0	0	0	4 (2)	NA

**Table 2** Prognostic assessment of the relative tumor-stroma cell subpopulation

Feature	Beta	HR (95% CI)	Wald	p value	Corrected p
T cells	-0.96	0.38 (0.207–0.703)	9.54	0.002	0.01
CD8 T cells	-0.06	0.94 (0.823–1.084)	0.66	0.415	0.415
Cytotoxic lymphocytes	-1.78	0.17 (0.066–0.426)	14.07	0.00018	0.002
B lineage	-0.57	0.57 (0.357–0.895)	5.93	0.015	0.037
NK cells	-0.86	0.42 (0.206–0.875)	5.39	0.02	0.04
Monocytic lineage	-0.75	0.47 (0.160–1.393)	1.85	0.174	0.222
Myeloid dendritic cells	-0.90	0.41 (0.218–0.755)	8.09	0.0045	0.015
Neutrophils	-2.26	0.10 (0.012–0.898)	4.23	0.04	0.066
Endothelial cells	-0.82	0.44 (0.107–1.810)	1.3	0.255	0.283
Fibroblasts	-1.07	0.34 (0.072–1.625)	1.82	0.177	0.222

survival analysis (HR=0.13,  $p=0.017$ ) (Additional file 1: Table S4).

Relative cytotoxic lymphocyte subpopulation abundance was used as the cellular TSH biomarker to separate patients into good- and poor-survival groups (threshold ratio, 0.3512). There was no significant difference in the cytotoxic lymphocyte subpopulation between the tumor and stromal tissues in the good-survival group (mean  $0.9233 \pm 0.36056$  vs.  $0.8628 \pm 0.2787$ , Wilcoxon signed rank  $p=0.455$ ), whereas in the poor-survival group, this cell subtype was significantly higher in the stroma than in the tumors ( $0.9631 \pm 0.2135$  vs.  $0.3531 \pm 0.1437$ , Wilcoxon signed rank  $p=7.755 \times 10^{-16}$ ) (Fig. 4a). This cell subpopulation was positively associated with good survival in breast cancer patients ( $p < 0.0001$ ) (Fig. 4b). This means that the tumor tissues in the good-survival group had higher levels of various cell subpopulations. Additionally, all cell subpopulation abundances were significantly lower in the tumor than in the stroma in the poor-survival group (Fig. 4c), whereas the opposite trend was observed in the good-survival group (Fig. 4d). A stronger correlation of cytotoxic lymphocyte cell subpopulation abundance between the tumor and stroma ( $r=0.45$ ,  $p=0.003$ ) was observed in the good-survival group, whereas a weaker correlation ( $r=0.11$ ,  $p=0.92$ ) was found in the poor-survival group (Additional file 1: Figure S1). The findings that the cytotoxic lymphocyte cell subpopulation is associated with survival were consistent with previous studies [46–48].

#### Analysis of tumor-stroma heterogeneity by cell subpopulation networks

Both the tumor and stroma exhibited distinct cell subpopulation networks in the good-survival and poor-survival groups, as depicted in Fig. 5a and b. Notably, the stroma-based cell subpopulation network displayed dense intrinsic connectivity, while the tumor-based

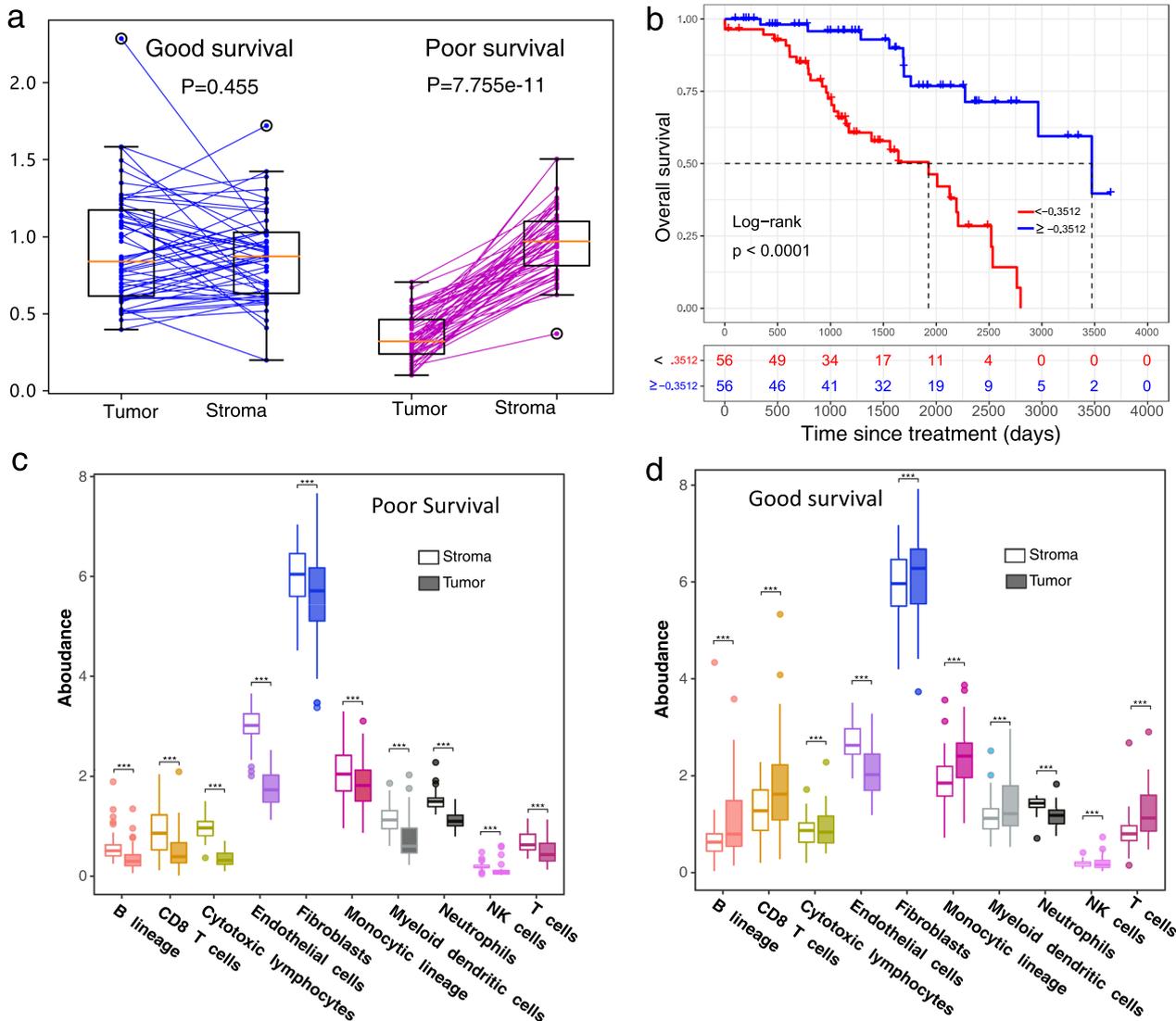
network exhibited sparse connectivity. Furthermore, the topological parameters ( $n=9$ ) of the cell subpopulation network differed significantly between the poor and good-survival groups for both stromal and tumor-based networks, as illustrated in Fig. 5c, d, and Additional file 1: Table S5.

In the stroma-based network, the good-survival group demonstrated a higher network diameter and greater node and edge numbers than the poor-survival group. Conversely, within the tumor-based network, the good-survival group exhibited a higher network edge number, density, and clustering coefficient than the poor-survival group. Interestingly, the good-survival group displayed relatively lower network centralization, indicating a greater dispersion of node centrality scores throughout the network in comparison to the maximum centrality score obtained. This finding suggests enhanced communication among cells, particularly T cells, to the central node in patients with poor survival.

#### Imaging cellular tumor-stroma heterogeneity

To establish a predictive model for the TSH score, we employed cell subpopulations derived from tumors. Figure 2 depicts a schematic representation of the radiogenomic analysis conducted on cellular TSH. Utilizing cross-validation-based feature selection, cytotoxic lymphocyte and fibroblast subpopulations were identified as key predictors targeting TSH, yielding an AUC of 0.962, with a sensitivity and specificity of 0.893 (Fig. 6a). To ascertain the reliability of our correlation map, we proceeded to validate the predictive model using an independent tumor-stroma dataset. Remarkably, the TSH score predicted by this model significantly stratified patients into distinct good and poor-survival groups ( $p=0.014$ ) (Fig. 6b).

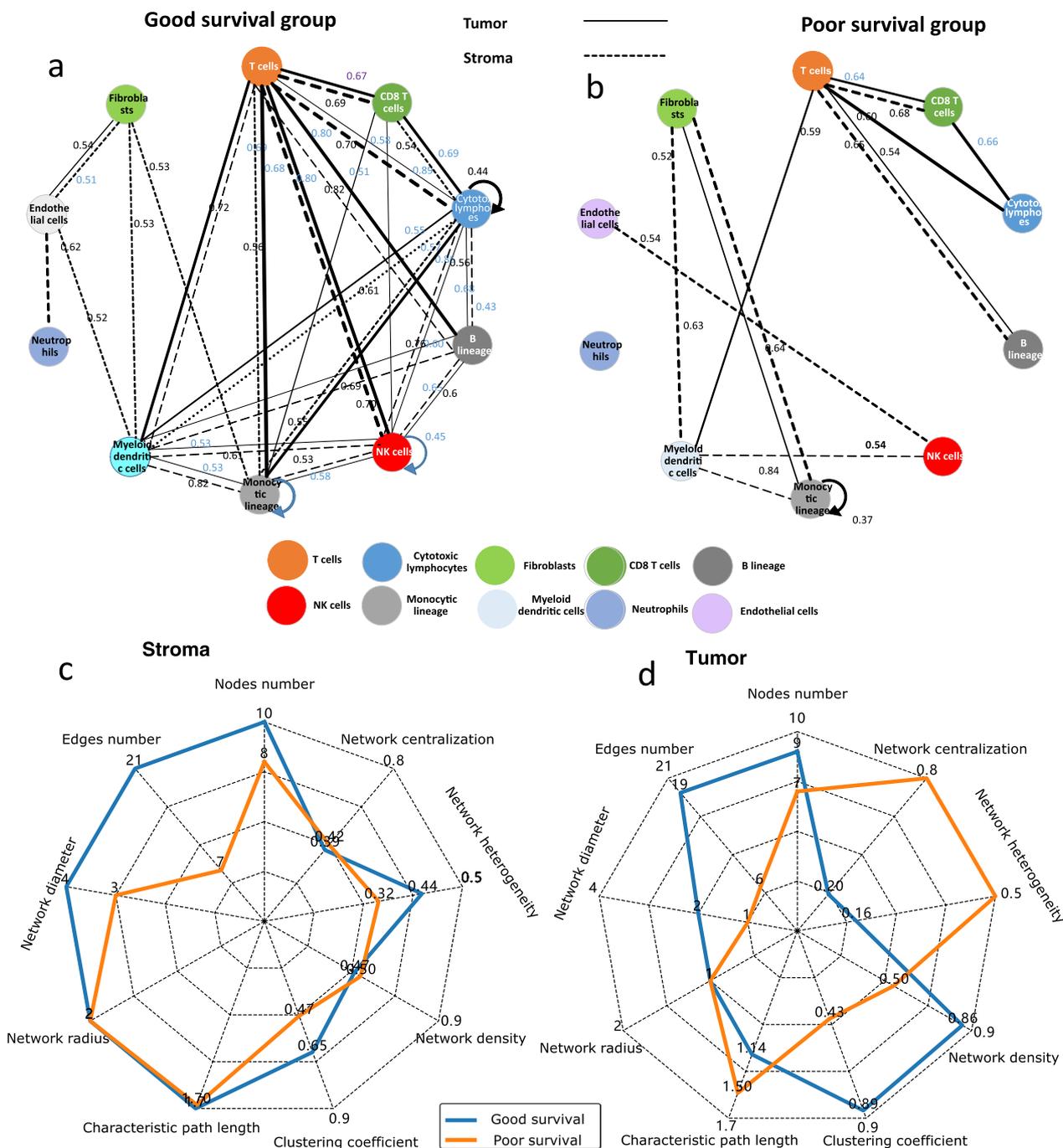
Subsequently, we applied the TSH model to the radiogenomic development dataset utilizing estimated



**Fig. 4** Distribution of cell subpopulations within and surrounding tumors in the good-survival and poor-survival groups. **a** The distributions of the abundance of cytotoxic lymphocyte cell subpopulations from tumor and stromal tissues in the good- and poor-survival groups. **b** Survival curves for the abundance of cytotoxic lymphocyte cell subpopulations. **c** Boxplots of cell subpopulation abundances in tumors and stromal regions in the poor-survival group. **d** Boxplots of cell subpopulation abundances in tumors and stromal regions in the good-survival group. In the boxplot, the centerline indicates the median; box limits indicate the 25% and 75% quantiles; whiskers represent the 1.5x interquartile range; and points above or below whiskers represent outliers

cell subpopulation data derived from tumors to predict the TSH score. Notably, the predicted TSH score successfully separated patients within the radiogenomic development dataset into distinct good-survival (n = 46) and poor-survival (n = 45) groups, using a threshold of 0.5. To establish an imaging feature-based predictive model for the classification of these groups based on the TSH score, we employed an optimal subset of imaging features, constituting the radiogenomic signature of TSH.

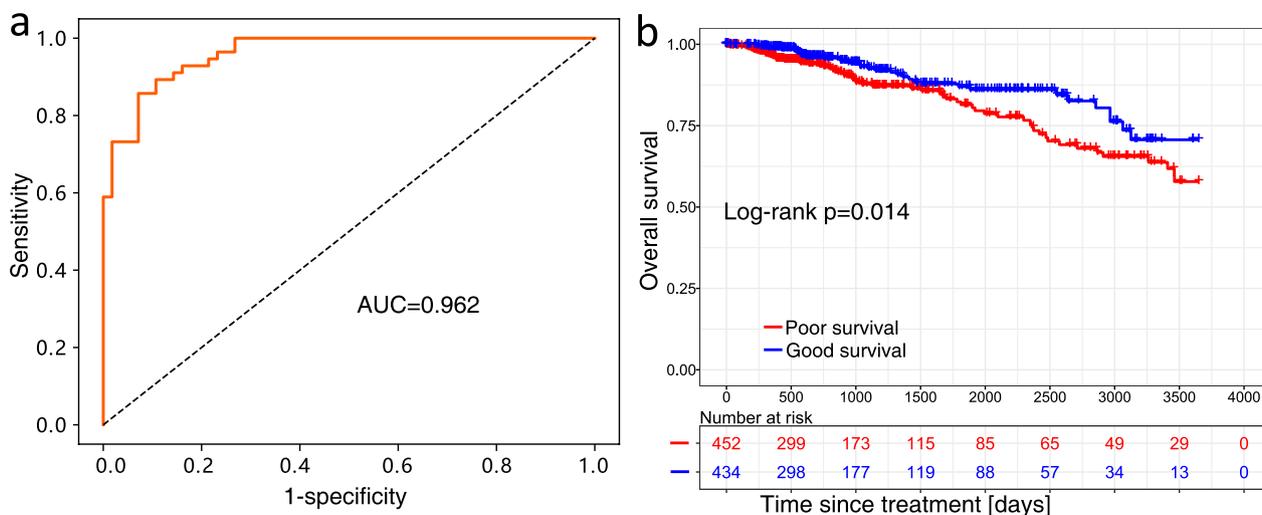
To identify this optimal subset of imaging features, we employed a genetic algorithm-based feature selection method with a population size of 100 and 100 iterations under tenfold cross-validation. Following feature selection, the predictive model retained six features, as outlined in Additional file 1: Table S6. Among these features, three were extracted from tumors, encompassing two morphological features (flatness and sphericity) and two texture features (long-run high gray level emphasis extracted from the precontrast image and inverse



**Fig. 5** Network topology in the poor-survival and good-survival groups. **a** The cell subpopulation network in the good-survival group; **b** The cell subpopulation network in the poor-survival group; **c** The radar map of the topological parameters (n=9) of the stroma-based cell subpopulation network for good- and poor-survival groups; **d** The radar map of the topological parameters (n=9) of the tumor-based cell subpopulation network for good- and poor-survival groups

difference moment normalized (IDMN) extracted from the postcontrast image). Additionally, two features were extracted from the stroma, including the IDMN feature and gray level nonuniformity (GLSZM).

The radiogenomic signatures identified in our study serve as surrogate markers for cellular TSH and hold significant prognostic value. These signatures provide informative assessments for patient prognosis, reflecting



**Fig. 6** Radiogenomic analysis and predictive model based on differences between the tumor and stroma. **a** The ROC curve for the predictive model. **b** The survival curves for the tumor-stromal validation dataset to evaluate the effectiveness of the identified relative cell subtype features

the potential of radiogenomics in enhancing our understanding of tumor-stroma interactions and facilitating personalized treatment decisions.

**Prognostic validation of the radiogenomic signatures**

The identified radiogenomic signatures underwent independent validation through survival analysis using two prognostic validation datasets: validation dataset 1 (n=180) and validation dataset 2 (n=61). Representative examples of the IDMN and flatness features are illustrated in Fig. 7a and b, respectively. The IDMN feature exhibited a significantly higher value ( $p = 6.386 \times 10^{-7}$ ) in the good-survival group than in the poor-survival group (prognostic validation 1 dataset, Fig. 7a). Moreover, elevated IDMN values or enhanced tumor flatness were indicative of a favorable prognosis in the prognostic validation 2 dataset (Additional file 1: Figure S2).

The radiogenomic signatures identified in this study were used to generate a predicted TSH score. Analysis of this score revealed a significant positive correlation with both overall survival (OS) ( $p = 0.038$ ) and recurrence-free survival (RFS) ( $p = 3 \times 10^{-4}$ ) in the prognostic validation 1 and prognostic validation 2 datasets, as illustrated in Fig. 7c and d, respectively. After adjusting for confounding factors such as age, HER2 status, ER status, PR status, and longest diameter, the radiogenomic signature maintained its significant prognostic value for OS ( $p = 0.004$ ).

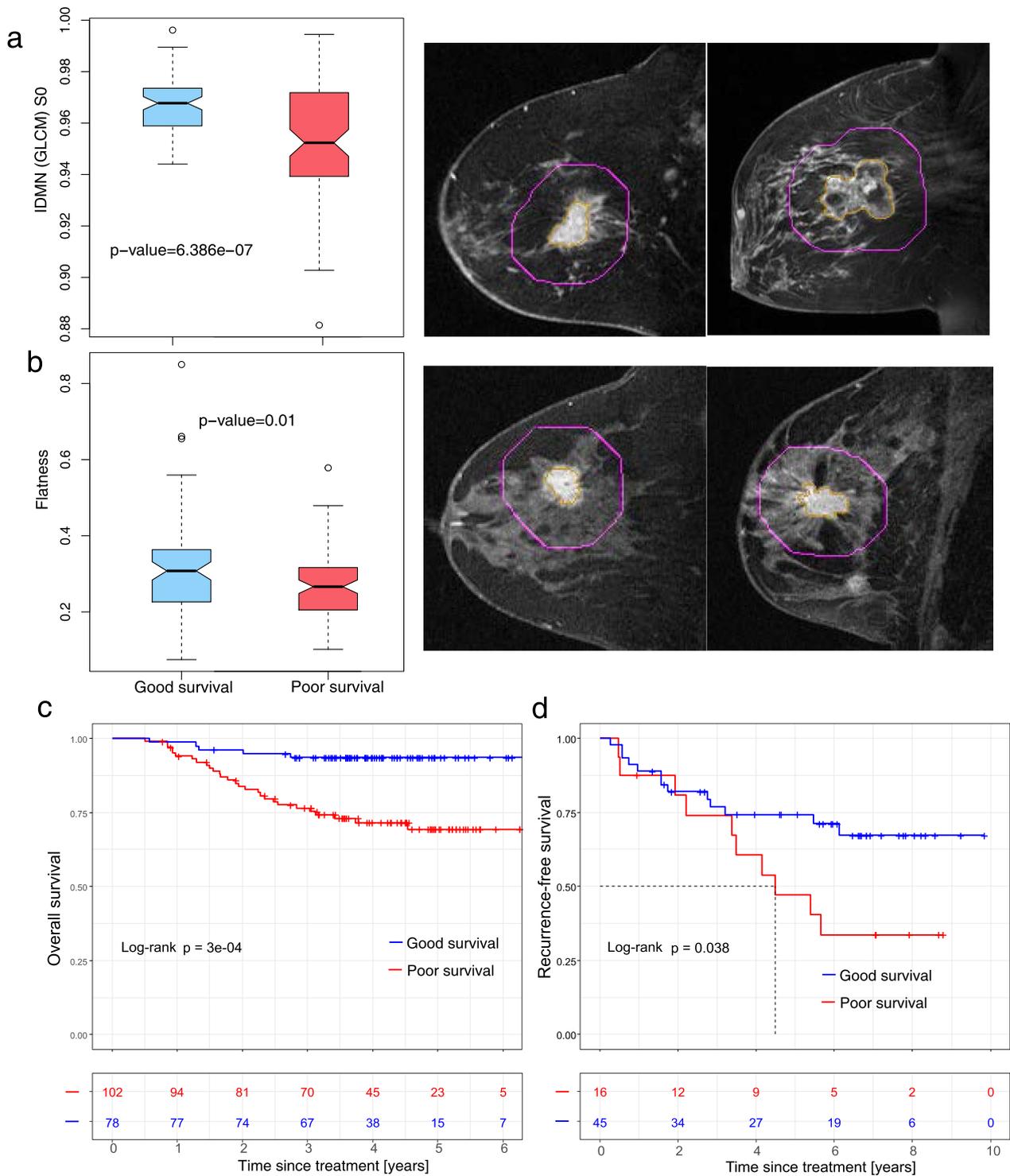
The validation of the identified radiogenomic signatures in two independent prognostic validation datasets strengthens their prognostic value. Therefore, the significant association between the predicted TSH score and both OS and RFS, coupled with its independent

prognostic value, highlights the relevance and reliability of the radiogenomic signatures identified in our study.

**Discussion**

In this multicohort study, we identified the radiogenomic signatures of TSH and independently evaluated their clinical implications. The cellular TSH biomarker was determined by assessing the relative differences in its abundance within the tumor and stromal cell subpopulations. These cellular TSH biomarkers demonstrated their effectiveness in differentiating patients into distinct groups with varying survival rates. To further validate the clinical significance of our findings, we developed a radiogenomic signature-based model targeting TSH and tested its predictive capabilities in two independent datasets. Overall, our study demonstrates the potential of radiogenomic analysis in uncovering hidden tumor phenotypes and highlights the promising role of these imaging-based surrogates in enhancing the accuracy of prognosis and treatment management.

We observed a positive correlation between cell subpopulation abundances estimated from the surrounding nontumoral tissue and poor survival, which is in line with previous findings [45]. In the network analysis of cell subpopulations, the good-survival group exhibited a denser interconnection among themselves compared to the poor-survival group, both in the tumor- and stroma-based cell subpopulation networks. A possible explanation is that an imbalanced microenvironment can influence tumor growth and progression in breast cancer [49–52]. The observed positive correlation between cell subpopulation abundances from the surrounding nontumoral tissue and poor OS, as well as the differential



**Fig. 7** Imaging feature distribution and Kaplan–Meier survival analyses. **a** Boxplot of the inverse difference moment normalized (IDMN) features in precontrast images showing significantly higher values in the good survival group than in the poor-survival group ( $p=6.386 \times 10^{-7}$ ). An example of a patient aged 54.16 years showed a higher feature value (0.9657) with good survival compared with a patient aged 44.5 years with poor survival (feature value = 0.9494). **b** Boxplot of the flatness feature showed significantly higher values in the good-survival group than in the poor-survival group ( $p=0.01$ ). Examples of a patient aged 62.4 years showed a higher feature value of 0.3086 with good survival than a patient aged 58.89 years with a feature value of 0.2700 with poor survival. The predicted tumor-stroma heterogeneity (TSH) score using radiogenomic signatures ( $n=6$ ) separated patients into good- and poor-survival groups for **c** overall survival (prognostic validation 1 dataset) and **d** recurrence-free survival (prognostic validation 2 dataset)

interconnectivity patterns within cell subpopulation networks, provides valuable insights into the complex microenvironment of breast cancer. Further investigations into the underlying mechanisms governing these observed network characteristics hold promise for identifying novel therapeutic strategies aimed at improving patient outcomes.

In our study, we observed a stronger correlation of cytotoxic lymphocyte cell subpopulation abundance between the tumor and stroma in the good-survival group, while a weaker correlation was found in the poor-survival group. Cytotoxic lymphocytes are essential in tumor immunity due to their correlation with tumor size, overall survival, immune checkpoint expression, and tumor microenvironment characteristics [46, 48]. The higher abundance of cytotoxic lymphocyte cell subpopulations in the tumor compared to the stroma can be attributed to enhanced tumor immune surveillance, increased tumor cell killing, activation of immune checkpoints, induction of antitumor immune memory, and modulation of the tumor microenvironment, all of which foster antitumor immune responses and inhibit tumor growth [53].

To comprehensively assess tumor-stroma phenotypes, we conducted an evaluation of radiogenomic signatures derived from both the tumor and surrounding parenchyma tissues. Specifically, higher tumor flatness and sphericity values are associated with poor survival, which is consistent with previous findings [54]. These features are related to a more irregular tumor shape or a higher level of image heterogeneity. Furthermore, we observed a significantly higher IDMN value in the stroma of the good-survival group than in the poor survival group. Our radiogenomic model employs radiogenomic signatures to estimate cellular TSH, which offers a non-invasive approach to evaluate the entire tumor and its surrounding stroma. This imaging-based evaluation can be repeated throughout the course of treatment, providing valuable longitudinal information. Importantly, our study demonstrates the feasibility of utilizing imaging signatures of TSH with biological significance, eliminating the need for genomic data from tumors or stromal tissues. By leveraging noninvasive imaging, we have the opportunity to provide valuable insights into the tumor-stroma interplay and its impact on patient outcomes, without the need for invasive biopsies at multiple sites.

Previous studies have investigated the association between radiomic features and the immune microenvironment in terms of TILs [26, 55]. Sun et al. developed a radiomic signature of CD8 T cells, which demonstrated an association with clinical response and outcomes in a cohort of 137 patients treated with anti-PD1 immunotherapy. Moreover, this signature was subsequently

employed to predict lesion response in 94 patients who underwent combined immunotherapy and radiotherapy [28]. A related study developed and validated a stromal imaging signature that could accurately diagnose and predict the survival benefit of adjuvant chemotherapy in resected gastric cancer [29]. Notably, our study differs from previous investigations, as we integrated imaging and matched cell subpopulation data from both within and outside the tumor and applied them to external datasets for survival analysis.

This study has some limitations. First, this was a retrospective study and may have been subject to patient selection bias. Second, the imaging data were collected from multiple cohorts, each with different imaging parameters. Therefore, it is crucial to conduct rigorous testing in future studies to assess the generalizability of the proposed imaging signatures in clinical applications. Third, the estimation of cell subpopulation abundance in this study relied on a deconvolution method utilizing mRNA data. It is worth noting that the accuracy of this method may impact the identification of cellular biomarkers and should be taken into consideration. Third, the data size is limited in this study. It is also important to note that acquiring both gene expression and imaging data simultaneously is particularly challenging and the availability of such data is limited. We have focused on obtaining paired gene expression and imaging data to train our models. This approach allows us to establish a correlation between the molecular characteristics represented by gene expression and the corresponding visual features observed through imaging techniques.

To address the critical challenge of incorporating prognostic information embedded in tumor-stroma heterogeneity, our predictive model utilizes novel radiogenomic signatures. These signatures capture the latent associations between the prognostic genomic signatures of TSH and radiomic features. As a result, these signatures have the potential to significantly enhance the accuracy of clinical prognostication with biological meanings. By providing valuable insights, our approach holds promise for improving patient management and guiding treatment decisions.

## Conclusions

We investigated cellular TSH and identified prognostic factors that reflect the environmental imbalance between the tumor and stroma. The relative abundance of cytotoxic lymphocyte subpopulations was used as a biomarker for cellular TSH, allowing for the stratification of patients into distinct survival groups. The good-survival group exhibited a more densely interconnected network of cell subpopulations, either within the tumor or stroma, than the poor-survival group. These cell biomarkers,

identified through radiological imaging signatures, have breast cancer-related biological implications and demonstrate prognostic power. The radiogenomic signature of cellular TSH was positively associated with good survival. External validation using multicohort datasets further supports their clinical relevance for prognostic assessments. This proposed model has the potential to refine prognosis analysis and guide personalized therapy for breast cancer patients who are likely to benefit from chemotherapy and have a favorable prognosis. Our framework could be extended to other cancer types to identify patients at high risk of recurrence. Further studies are required to confirm the clinical implications and establish the clinical utility of this approach.

#### Abbreviations

TSH	Tumor-stroma heterogeneity
AUC	Area under the ROC curve
TILs	Tumor-infiltrating lymphocytes
DCE-MRI	Dynamic contrast-enhanced magnetic resonance imaging
HR	Hazard ratio
OS	Overall survival
RFS	Recurrence-free survival

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-023-04748-6>.

**Additional file 1: Figure S1.** Relative values of the cytotoxic lymphocyte cell subpopulations. **Figure S2.** Distributions of imaging features in the good survival and poor survival groups. Significantly higher a) inverse difference moment normalized (IDMN) feature values in precontrast images and b) significantly higher tumor flatness values in the good survival group than in the poor survival group. **Table S1.** Imaging feature list. **Table S2.** Tumor cell subpopulations associated with survival in breast cancer. **Table S3.** Stromal cell subpopulations associated with survival in breast cancer. **Table S4.** Multivariate analysis of tumor/stroma subpopulations associated with survival in breast cancer. **Table S5.** Network characteristics of cell-to-cell connections. **Table S6.** Radiogenomic signatures in the predictive model.

#### Author contributions

MF and LL designed the study, analyzed the data and wrote the manuscript. YZ, KW, YH and YG contributed to the study design, image processing, gene expression data analysis and statistical analysis. ZL participated in the acquisition, annotation and curation of the data. All the authors have read and approved the final manuscript.

#### Funding

This work was supported by the National Natural Science Foundation of China (62271178, U21A20521) and the Natural Science Foundation of Zhejiang Province of China (LR23F010002).

#### Availability of data and codes

The gene expression data of tumor-stroma development and tumor-stroma validation datasets are available from the TCGA-BRCA project of Genomic Data Commons. [<https://portal.gdc.cancer.gov/projects/TCGA-BRCA>]. The imaging data of the tumor-stroma development and tumor-stroma validation datasets are available from TCGA-BRCA in TCIA on the website. [<https://wiki.cancerimagingarchive.net/display/Public/TCGA-BRCA>]. Breast-MRI-NACT-Pilot is available in TCIA on the website. [<https://wiki.cancerimagingarchive.net/display/Public/Breast-MRI-NACT-Pilot>]. The ISPY-1 trial is available in TCIA on the website. [<https://wiki.cancerimagingarchive.net/display/Public/ISPY1>].

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the institutional review board of Hangzhou Dianzi University, and informed consent was waived owing to the retrospective nature of this study.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 21 July 2023 Accepted: 22 November 2023

Published online: 25 November 2023

#### References

1. Polyak K. Heterogeneity in breast cancer. *J Clin Invest.* 2011;121:3786–8.
2. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med.* 2013;19:1423–37.
3. Valkenburg KC, de Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. *Nat Rev Clin Oncol.* 2018;15:366–81.
4. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol.* 2020;30:R921–5.
5. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell.* 2012;21:309–22.
6. Joyce JA, Fearon DT. T cell exclusion, immune privilege, and the tumor microenvironment. *Science.* 2015;348:74–80.
7. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med.* 2018;24:541–50.
8. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, Budczies J, Huober J, Klauschen F, Furlanetto J, et al. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. *Lancet Oncol.* 2018;19:40–50.
9. McDonald KA, Kawaguchi T, Qi Q, Peng X, Asaoka M, Young J, Opyrchal M, Yan L, Patnaik S, Otsuji E, Takabe K. Tumor heterogeneity correlates with less immune response and worse survival in breast cancer patients. *Ann Surg Oncol.* 2019;26:2191–9.
10. Gu-Trantien C, Loi S, Garaud S, Equeter C, Libin M, de Wind A, Ravoet M, Le Buanec H, Sibille C, Manfouo-Foutsop G, et al. CD4(+) follicular helper T cell infiltration predicts breast cancer survival. *J Clin Invest.* 2013;123:2873–92.
11. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, Bassez A, Decaluwe H, Pircher A, Van den Eynde K, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med.* 2018;24:1277–89.
12. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12:453–7.
13. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, Selves J, Laurent-Puig P, Sautes-Fridman C, Fridman WH, de Reynies A. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17:218.
14. Tekpli X, Lien T, Rossevoold AH, Nebdal D, Borgen E, Ohnstad HO, Kyte JA, Vallon-Christersson J, Fongaard M, Due EU, et al. An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment. *Nat Commun.* 2019;10:5499.
15. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelshen J, Orntberg L, Gevaert O, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell.* 2018;173(338–354): e315.

16. Mahal BA, Alshalalfa M, Zhao SG, Beltran H, Chen WS, Chipidza F, Davicioni E, Karnes RJ, Ku SY, Lotan TL, et al. Genomic and clinical characterization of stromal infiltration markers in prostate cancer. *Cancer*. 2020;126:1407–12.
17. Jiang L, You C, Xiao Y, Wang H, Su GH, Xia BQ, Zheng RC, Zhang DD, Jiang YZ, Gu YJ, Shao ZM. Radiogenomic analysis reveals tumor heterogeneity of triple-negative breast cancer. *Cell Rep Med*. 2022;3: 100694.
18. Li H, Zhu Y, Burnside ES, Huang E, Drukker K, Hoadley KA, Fan C, Conzen SD, Zuley M, Net JM, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TCIA data set. *NPJ Breast Cancer*. 2016;2:16012.
19. 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.
20. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, Conzen SD, Whitman GJ, Sutton EJ, Net JM, et al. MR Imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of mammprint, oncotype DX, and PAM50 gene assays. *Radiology*. 2016;281:382–91.
21. Bommi JR, Kummari S, Lakavath K, Sukumaran RA, Panicker LR, Marty JL, Goud KY. Recent trends in biosensing and diagnostic methods for novel cancer biomarkers. *Biosensors-Basel*. 2023;13:398.
22. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. *They Are Data Radiol*. 2016;278:563–77.
23. 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.
24. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue R, Even AJG, Jochems A, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–62.
25. Panayides AS, Pattichis MS, Leandrou S, Pitris C, Constantinidou A, Pattichis CS. Radiogenomics for precision medicine with a big data analytics perspective. *IEEE J Biomed Health Inform*. 2019;23:2063–79.
26. Wu J, Li X, Teng X, Rubin DL, Napel S, Daniel BL, Li R. Magnetic resonance imaging and molecular features associated with tumor-infiltrating lymphocytes in breast cancer. *Breast Cancer Res*. 2018;20:101.
27. Fan M, Xia P, Clarke R, Wang Y, Li L. Radiogenomic signatures reveal multi-scale intratumour heterogeneity associated with biological functions and survival in breast cancer. *Nat Commun*. 2020;11:4861.
28. 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.
29. Jiang Y, Liang X, Han Z, Wang W, Xi S, Li T, Chen C, Yuan Q, Li N, Yu J, et al. Radiographical assessment of tumour stroma and treatment outcomes using deep learning: a retrospective, multicohort study. *Lancet Digit Health*. 2021;3:e371–82.
30. Kim JH, Ko ES, Lim Y, Lee KS, Han BK, Ko EY, Hahn SY, Nam SJ. Breast cancer heterogeneity: mr imaging texture analysis and survival outcomes. *Radiology*. 2017;282:665–75.
31. Wandler F, Favicchio R, Simon T, Alifrangis C, Stebbing J, Giamas G. Extracellular vesicles swarm the cancer microenvironment: from tumor-stroma communication to drug intervention. *Oncogene*. 2017;36:877–84.
32. Lyssiotis CA, Kimmelman AC. Metabolic interactions in the tumor microenvironment. *Trends Cell Biol*. 2017;27:863–75.
33. Ashworth A, Lord CJ, Reis-Filho JS. Genetic interactions in cancer progression and treatment. *Cell*. 2011;145:30–8.
34. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Cancer Res*. 2019;79:4557–66.
35. Garcia-Martinez E, Gil GL, Benito AC, Gonzalez-Billalabeitia E, Conesa MA, Garcia Garcia T, Garcia-Garre E, Vicente V, Ayala de la Pena F. Tumor-infiltrating immune cell profiles and their change after neoadjuvant chemotherapy predict response and prognosis of breast cancer. *Breast Cancer Res*. 2014;16:488.
36. Danenberg E, Bardwell H, Zanotelli VRT, Provenzano E, Chin SF, Rueda OM, Green A, Rakha E, Aparicio S, Ellis IO, et al. Breast tumor microenvironment structures are associated with genomic features and clinical outcome. *Nat Genet*. 2022;54:660–9.
37. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61–70.
38. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045–57.
39. Li KL, Partridge SC, Joe BN, Gibbs JE, Lu Y, Esserman LJ, Hylton NM. Invasive breast cancer: predicting disease recurrence by using high-spatial-resolution signal enhancement ratio imaging. *Radiology*. 2008;248:79–87.
40. Hylton NM, Gatsonis CA, Rosen MA, Lehman CD, Newitt DC, Partridge SC, Bernreuter WK, Pisano ED, Morris EA, Weatherall PT, et al. Neoadjuvant chemotherapy for breast cancer: functional tumor volume by MR imaging predicts recurrence-free survival—results from the ACRIN 6657/CALGB 150007 I-SPY 1 TRIAL. *Radiology*. 2016;279:44–55.
41. Smith K. Single site breast DCE-MRI data and segmentations from patients undergoing neoadjuvant chemotherapy. The Cancer Imaging Archive. 2016.
42. Fan M, Zhang P, Wang Y, Peng W, Wang S, Gao X, Xu M, Li L. Radiomic analysis of imaging heterogeneity in tumours and the surrounding parenchyma based on unsupervised decomposition of DCE-MRI for predicting molecular subtypes of breast cancer. *Eur Radiol*. 2019;29:4456–67.
43. Fan M, Yuan W, Zhao W, Xu M, Wang S, Gao X, Li L. Joint prediction of breast cancer histological grade and Ki-67 expression level based on DCE-MRI and DWI radiomics. *IEEE J Biomed Health Inform*. 2020;24:1632–42.
44. 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.
45. Goff SL, Danforth DN. The role of immune cells in breast tissue and immunotherapy for the treatment of breast cancer. *Clin Breast Cancer*. 2021;21:e63–73.
46. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J Clin Oncol*. 2011;29:1949–55.
47. Zhang S, Wang J, Zhang X, Zhou F. Tumor-infiltrating CD8+ lymphocytes predict different clinical outcomes in organ- and non-organ-confined urothelial carcinoma of the bladder following radical cystectomy. *PeerJ*. 2017;5: e3921.
48. Han Y, Wang J, Xu B. Cytotoxic lymphocyte-related gene signature in triple-negative breast cancer. *J Pers Med*. 2023;13:457.
49. Jesser EA, Brady NJ, Huggins DN, Witschen PM, O'Connor CH, Schwertfeger KL. STAT5 is activated in macrophages by breast cancer cell-derived factors and regulates macrophage function in the tumor microenvironment. *Breast Cancer Res*. 2021;23:104.
50. Vathiotis IA, Moutafi MK, Divakar P, Aung TN, Qing T, Fernandez A, Yaghoobi V, El-Abed S, Wang Y, Guillaume S, et al. Alpha-smooth muscle actin expression in the stroma predicts resistance to trastuzumab in patients with early-stage HER2-positive breast cancer. *Clin Cancer Res*. 2021;27:6156–63.
51. Chandra Jena B, Sarkar S, Rout L, Mandal M. The transformation of cancer-associated fibroblasts: Current perspectives on the role of TGF-beta in CAF mediated tumor progression and therapeutic resistance. *Cancer Lett*. 2021;520:222–32.
52. Ibrahim AM, Bilsland A, Rickelt S, Morris JS, Stein T. A matresome RNA signature from early-pregnancy mouse mammary fibroblasts predicts distant metastasis-free breast cancer survival in humans. *Breast Cancer Res*. 2021;23:90.
53. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoeediting: from immunosurveillance to tumor escape. *Nat Immunol*. 2002;3:991–8.
54. Fan M, Cui Y, You C, Liu L, Gu Y, Peng W, Bai Q, Gao X, Li L. Radiogenomic signatures of oncotype DX recurrence score enable prediction of survival in estrogen receptor-positive breast cancer: a multicohort study. *Radiology*. 2022;302:516–24.
55. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, Corredor G, Alilou M, Bera K, Fu P, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer Immunol Res*. 2020;8:108–19.

## Publisher's Note

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