



Co-clinical FDG-PET radiomic signature in predicting response to neoadjuvant chemotherapy in triple-negative breast cancer

Sudipta Roy¹ · Timothy D. Whitehead¹ · Shunqiang Li² · Foluso O. Ademuyiwa² · Richard L. Wahl^{1,3} · Farrokh Dehdashti¹ · Kooresh I. Shoghi^{1,4}

Received: 27 April 2021 / Accepted: 4 July 2021 / Published online: 30 July 2021
© The Author(s) 2021, corrected publication 2021

Abstract

Purpose We sought to exploit the heterogeneity afforded by patient-derived tumor xenografts (PDX) to first, optimize and identify robust radiomic features to predict response to therapy in subtype-matched triple negative breast cancer (TNBC) PDX, and second, to implement PDX-optimized image features in a TNBC co-clinical study to predict response to therapy using machine learning (ML) algorithms.

Methods TNBC patients and subtype-matched PDX were recruited into a co-clinical FDG-PET imaging trial to predict response to therapy. One hundred thirty-one imaging features were extracted from PDX and human-segmented tumors. Robust image features were identified based on reproducibility, cross-correlation, and volume independence. A rank importance of predictors using ReliefF was used to identify predictive radiomic features in the preclinical PDX trial in conjunction with ML algorithms: classification and regression tree (CART), Naïve Bayes (NB), and support vector machines (SVM). The top four PDX-optimized image features, defined as radiomic signatures (RadSig), from each task were then used to predict or assess response to therapy. Performance of RadSig in predicting/assessing response was compared to SUV_{mean} , SUV_{max} , and lean body mass-normalized SUL_{peak} measures.

Results Sixty-four out of 131 preclinical imaging features were identified as robust. NB-RadSig performed highest in predicting and assessing response to therapy in the preclinical PDX trial. In the clinical study, the performance of SVM-RadSig and NB-RadSig to predict and assess response was practically identical and superior to SUV_{mean} , SUV_{max} , and SUL_{peak} measures.

Conclusions We optimized robust FDG-PET radiomic signatures (RadSig) to predict and assess response to therapy in the context of a co-clinical imaging trial.

Keywords Triple-negative breast cancer (TNBC) · FDG-PET · Radiomics · Co-clinical imaging · Quantitative imaging · Machine learning

Introduction

Triple-negative breast cancer (TNBC) is a highly heterogeneous and aggressive cancer characterized by poor outcome and higher relapse rates compared to other subtypes of breast cancer. Pathological complete response (pCR) is often used as a critical endpoint in the treatment of TNBC following neoadjuvant chemotherapy (NAC) as it is often associated with favorable long-term outcome. Therefore, it is critical to identify patients who will respond to NAC therapy to avoid the use of ineffective treatments. Intratumoral heterogeneity is regarded as a major factor in tumor progression and resistance to NAC [1]. Towards that end, advanced quantitative imaging (QI) strategies, including extraction of image features, or radiomics, have

This article is part of the Topical Collection on Oncology - General.

✉ Kooresh I. Shoghi
shoghik@wustl.edu

¹ Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

² Department of Medicine, Division of Oncology, Washington University School of Medicine, St. Louis, MO, USA

³ Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO, USA

⁴ Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA

been employed to characterize tumor heterogeneity and to predict/assess response to therapy [2, 3].

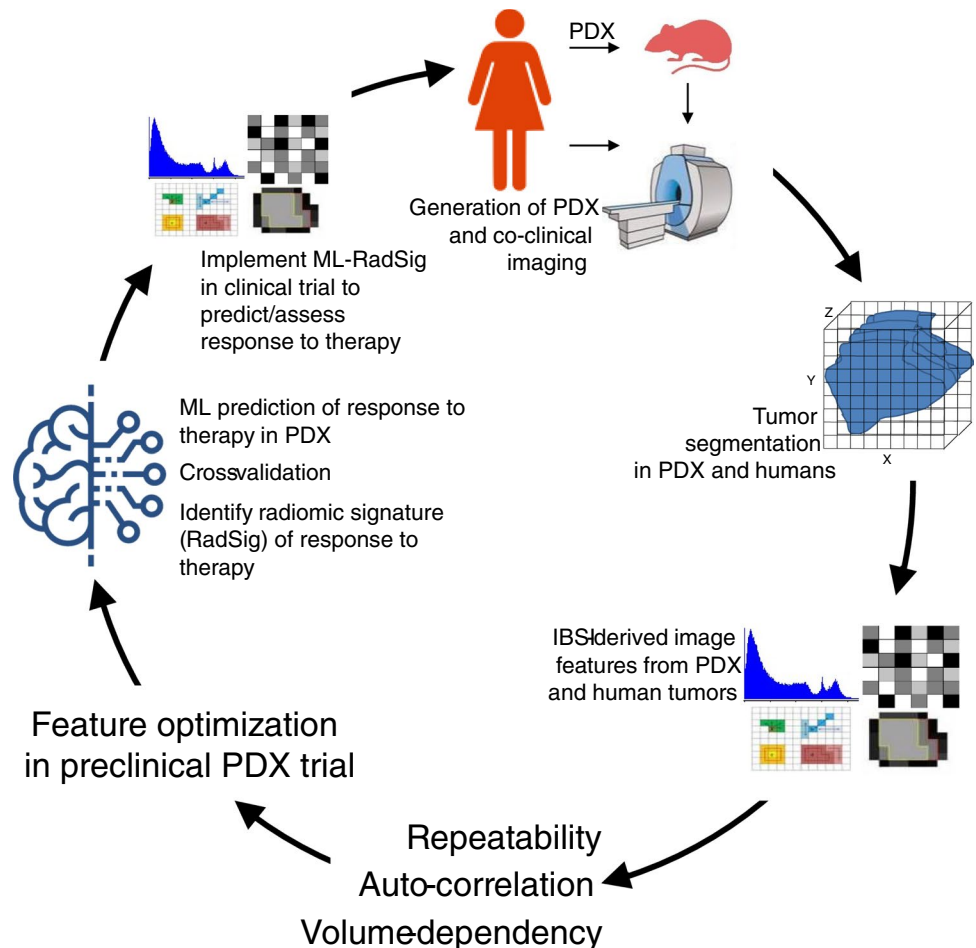
We designed a co-clinical trial to assess the efficacy of docetaxel/carboplatin therapy in patients with TNBC and patient-derived tumor xenografts (PDX) generated from TNBC patient biopsies. Co-clinical trials are an emerging area of investigation in which a clinical trial is coupled with a corresponding (subtype-matched or patient-specific) preclinical trial to inform the corresponding clinical trial [4–10]. The emergence of PDXs as co-clinical platforms is largely motivated by the realization that established cell lines do not recapitulate the heterogeneity of human tumors and the diversity of tumor phenotypes [11]. Indeed, numerous investigations have demonstrated that PDX accurately reflect patients’ tumors in terms of the histomorphology, gene expression profiles, and gene copy number alterations [12–16], as well as the ability to predict therapeutic response in patients, especially when a clinically relevant drug dosage is used [17–19]. To that end, the National Cancer Institute’s (NCI) Patient-Derived Models Repository (<https://pdmr.cancer.gov>), EuroPDX (<https://www.europdx.eu>), academic institutions, and numerous commercial entities have launched wide-ranging

PDX repositories to advance the use of PDX in precision medicine.

One of the objectives of the co-clinical trial, which is still underway, is to predict response to therapy using [¹⁸F] fluorodeoxyglucose (FDG) with positron emission tomography (PET). We previously identified six TNBC subtypes including 2 basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), a mesenchymal stem-like (MSL), and a luminal androgen receptor (LAR) subtype through molecular signatures of TNBC subtypes [20]. The use of PDX in preclinical imaging offers numerous advantages in translational imaging research, chief among them is retention of human tumor heterogeneity [12, 16, 21], which can be exploited to develop image metrics of response to therapy. Thus, the objective of this work was to first, optimize and identify robust radiomic features to predict response to therapy in subtype-matched TNBC PDX, and second, implement PDX-optimized image features in the TNBC co-clinical study to predict response to therapy using machine learning (ML) algorithms.

The scheme outlined in Fig. 1 highlights the paradigm we undertook in this effort. We used the co-clinical imaging trial to define, for the first time, parallels in radiomic

Fig. 1 Overview of methodology in optimizing radiomic features in the co-clinical trial. TNBC PDX were generated from human tumor biopsies. Tumors were segmented following co-clinical imaging to extract radiomic features. Radiomic features were extracted per IBSI guidelines. Repeatability, cross-correlation, and volume dependency were performed to identify the robust features. ReliefF and then ML were used to predict/assess the response to therapy in PDX and to identify radiomic signatures (RadSig). RadSig was implemented in the clinical trial to predict/assess response to therapy



features between preclinical and clinical imaging. To address the primary objective, we characterized the reproducibility, cross-correlation (auto-correlation), and volume dependency of FDG-PET radiomic features in PDX. Optimal radiomic features were then used in ML algorithms to define radiomic signatures (RadSig) of response to therapy in the preclinical PDX trial. With the RadSig at hand, we used RadSig to predict response to therapy in the preclinical arm. To address the secondary objective, we performed an interim analysis to implement radiomic signatures optimized in the preclinical PDX trial to predict response to therapy in the clinical arm. Our findings suggest that RadSig performed significantly better than SUV measures to predict (using baseline metrics) and assess (difference in image metrics) response to therapy in both preclinical and clinical arms.

Methods

Co-clinical protocol

The co-clinical design is outlined in the scheme of Fig. 2A and described below. Twenty newly diagnosed stage II or III TNBC patients were recruited into an ongoing co-clinical trial. Patient inclusion and exclusion selection criteria are detailed in ClinicalTrials.gov ID # NCT02124902. A secondary goal of the co-clinical trial was to assess the performance of FDG-PET in predicting/assessing response to therapy. TNBC PDX were generated as previously described [22] from TNBC patient tumor repository. Briefly, 6- to 10-week-old female NOD scid gamma (NSG) mice were obtained from The Jacksons Laboratory (<https://www.jax.org>). Mice

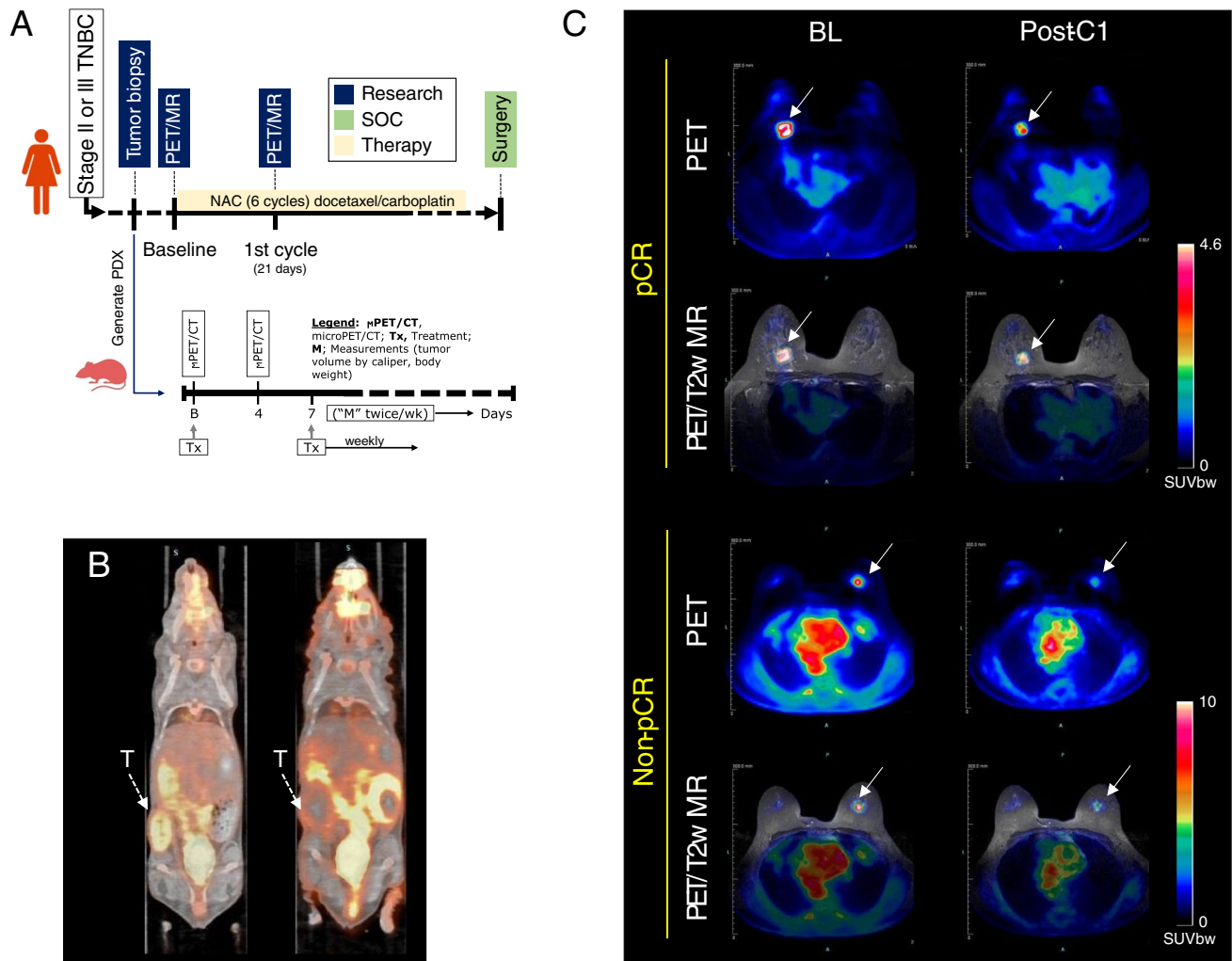


Fig. 2 Co-clinical imaging. **A** Co-clinical imaging study design. Stage II or III TNBC patients were recruited into the study for 6 cycles of docetaxel/carboplatin therapy. Imaging timepoints are indicated on the timeline. PDX are generated from patient tumor biop-

ties to assess response to therapy with imaging at baseline and 4 days post-therapy. **B** Representative preclinical PET/CT and **C** clinical PET/MR images of response to therapy. Tumor is indicated by a white arrow

were anesthetized with isoflurane and an inverted Y-shaped incision was made along the thoracic-inguinal region to expose the 4th inguinal mammary fat pad. Two to four million tumor cells mixed with Matrigel in a volume of 30 μ l were injected into the mammary fat pad. Following engraftment, tumor growth in PDX mice was monitored for recruitment into the preclinical trial. TNBC PDX subtypes were identified as described previously [23] based on molecular signature analysis of 93 TNBC PDX to identify TNBC subtypes including basal-like (BL1 and BL2), an immunomodulatory (IM), a mesenchymal (M), and a luminal androgen receptor (LAR) subtype.

Preclinical imaging Small animal PET/CT was performed on the Inveon microPET/CT scanner as described previously [23]. Briefly, 4 h prior to imaging session, food was removed from cages while water was given ad libitum. Mice were anesthetized with 2–2.5% isoflurane by inhalation via an induction chamber. Anesthesia was maintained throughout the imaging session by delivering 1–1.5% isoflurane via a custom-designed nose cone. A heat lamp was used to maintain body temperature. TNBC PDX were injected with ¹⁸F-FDG (6.66–8.14 MBq) by tail vein immediately before a 0–60-min dynamic small animal PET acquisition. Images were reconstructed with a 3D OSEM algorithm with a ramp filter at 0.5 cutoff and voxel size of \sim 0.8 mm isotropic.

Preclinical therapeutic studies TNBC PDX ($N=29$) were imaged at baseline (BL) and 4 days (4D) following start of therapy (Fig. 2A). Docetaxel (20 mg/kg IP)/carboplatin (50 mg/kg IP) was administered following BL imaging and weekly for a period of 4 weeks. Tumor volumes were measured bi-weekly using the formula $\text{volume} = 1/6 * L * W^2$ where L and W represent the length and width of the tumor, respectively. All animal experiments were conducted in compliance with the Guidelines for the Care and Use of Research Animals established by Washington University's Animal Studies Committee.

Clinical imaging Simultaneous FDG-PET and MR imaging protocol was implemented on the Siemens Biograph mMR. Subjects were imaged at baseline (BL) prior to therapy and between the first cycle (C1) and second cycle of docetaxel/carboplatin for a total of 6 cycles (21 days per cycle). At each imaging time point, patients were fasted for \sim 4 h prior to injection of \sim 10 mCi of FDG. After an uptake period, patients were positioned prone on the PET/MR scanner. FDG-PET imaging was performed starting at 30 to 70 min post FDG administration for a total of 40 min acquisition to accommodate the simultaneous MR acquisition protocol. Default Dixon sequence was used for attenuation correction. Images were reconstructed to produce four 10-min frames.

In parallel with FDG-PET acquisition, T1-weighted (T1w) and T2-weighted (T2w) MR acquisitions were performed.

Image analysis and extraction of radiomic features

Preclinical imaging Static 10-min PET/CT images obtained 50-min post-administration of FDG (representative image in Fig. 2B) were processed in two steps. In the first step, co-registered PET/CT images were analyzed using the Inveon Research Workplace (IRW) software (Siemens Healthcare). Volumes of interest (VOIs) were manually drawn on co-registered PET/CT images to include tumor(s). Second, VOIs and individual voxels were normalized to SUV in MATLAB using the relation: $\text{SUV} = [\text{activity (Bq/mL)}] \times [\text{animal weight (g)}] / [\text{injected dose (Bq)}]$.

Clinical imaging Tumor VOIs were manually drawn on 20-min static PET images obtained by averaging two 10-min frames 50–70-min post-administration of FDG (representative image in Fig. 2C). To ensure harmonization of preclinical and clinical pipelines, IRW was used to segment tumors on PET/MR images. Mean SUV (SUV_{mean}) for the entire tumor was calculated as per above. SUV_{max} was determined by identifying the maximum voxel activity in the tumor VOI. Peak SUV was normalized to lean body mass (SUL_{peak}) based on positron emission tomography response criteria in solid tumors (PERCIST) [24].

Extraction of imaging features One hundred thirty-one imaging features were extracted from preclinical and clinical tumors. These include one hundred twenty radiomic features, tumor volume, metabolic tumor volume, and nine SUV metrics as tabulated in Supplemental Table S1. Radiomic features were determined per the image biomarker standardization initiative (IBSI) guidelines [25, 26]. Equal-probability quantization algorithms to quantize raw data into gray level (Ng) were implemented using `histeq` MATLAB function. Resampling to isotropic voxel size in all three directions was applied to all higher order features. Thirty-seven first-order features were extracted directly from raw data. All higher order features were extracted after applying fixed quantization of gray level $\text{Ng} = 64$.

Robustness of radiomic features

We evaluated the robustness of radiomic features in terms of reproducibility (test–retest), cross-correlation, and the dependency on tumor volume. Robust radiomic features were then used as predictors of response to therapy.

Test–retest A preclinical test–retest protocol was implemented to optimize the reproducibility of radiomic features.

PDX ($N=40$) were imaged on consecutive days (day 1 and day 2) in identical conditions.

Cross-correlation The cross-correlation between features was determined using Spearman correlation. A threshold Spearman correlation of $\rho \geq 0.9$ and significance value $P < 0.001$ were chosen to signify high correlation between features.

Volume-dependent radiomic features Radiomic features were regressed against their corresponding tumor volumes. Linear or nonlinear functional forms were used to fit all significant volume-dependent features.

Prediction and assessment of response to therapy

Prediction vs. assessment of response to therapy We make a distinction between predicting and assessing response to therapy. In predicting response to therapy, BL imaging features were used to predict response to therapy in either the preclinical or the clinical arm. In assessing response to therapy, the change (Δ) in image feature between on-treatment (4 days post baseline imaging in preclinical and post-C1 in clinical) and BL was used to predict response to therapy in the preclinical or the clinical arms.

Classification of response to therapy In preclinical studies, endpoint caliper volume change from start of treatment was considered as surrogate of response to therapy with response to therapy corresponding to $> 20\%$ decrease in volume, partial response corresponding to $\leq 20\%$ change in volume, and no response corresponding to $> 20\%$ increase in volume. Baseline radiomic features and change in radiomic features between 4d post-treatment and baseline scans were used as the predictive criterion for ML algorithms. In clinical studies, pCR was used to determine response to therapy. pCR was defined as no histological evidence of invasive tumor cells in the surgical breast specimen and sentinel or axillary lymph nodes.

Feature selection In preclinical studies, the relief-based algorithm (RBA) [27] was used to select a subset of features as inputs to the ML algorithms. A relevance threshold ($\tau=0.05$) [28] was used to select most relevant weighted features to facilitate in expansive modeling, reduce overfitting, and make the task tractable for inputs in ML algorithms. These optimal features were used to predict response or assess response to therapy using BL and difference between on-treatment and BL optimal features, respectively.

Machine learning for outcome prediction The ML algorithms used in this study include CART [29], SVM [30],

and NB [31]. In implementing CART, Gini index was used at each partition to determine splitting criteria with a binary threshold of CART. In implementing SVM, radial basis function (RBF) kernel was used to make the hyperplane decision boundary between the classes. Objective function L2-norm regularization was used to overcome overfitting problem. CART, SVM, and NB work well with datasets as low as $N=20$ [32]. Ten-fold cross-validation was used to avoid overfitting the ML model [33].

Statistical analyses

Robustness of features Lin's concordance correlation coefficient (LCC) [34] was used to assess reproducibility using Stata version 12.1. $LCC \geq 0.7$ was considered as a threshold of reproducible radiomic feature [35, 36]. As indicated above, cross-correlation between features was evaluated using the Spearman correlation $\rho \geq 0.9$ at significance value $P < 0.001$. To display clusters of correlations, hierarchical clustering of the Spearman correlation heatmap was performed. In evaluating volume dependency of features, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were calculated for each functional form, and the appropriate model was selected based on the minimum value of AIC and BIC. The Spearman correlation (ρ) was used to determine the correlation between each feature and tumor volume.

Performance metrics of response to therapy prediction Common performance metrics including accuracy, F-score, sensitivity, specificity, precision, and negative predictive value (NPV) were used to assess performance of response to therapy [20]. The performance of the radiomic features was additionally compared with SUV_{mean} , SUV_{max} , and SUL_{peak} based on PERCIST [24].

Results

Reproducibility of preclinical radiomic features

Test–retest was performed to assess the reproducibility of radiomic features using LCC as a measure of reproducibility. Ninety-four out of 129 radiomic features (72.9%) were identified as reproducible with $LCC \geq 0.7$. The frequency of correlations along with the cumulative percent is displayed in Fig. 3A. Approximately 22% of features were highly reproducible with $LCC \geq 0.9$. The reproducibility by class of features is depicted in Fig. 3B. Figure 3C depicts the LCC values of all reproducible radiomic

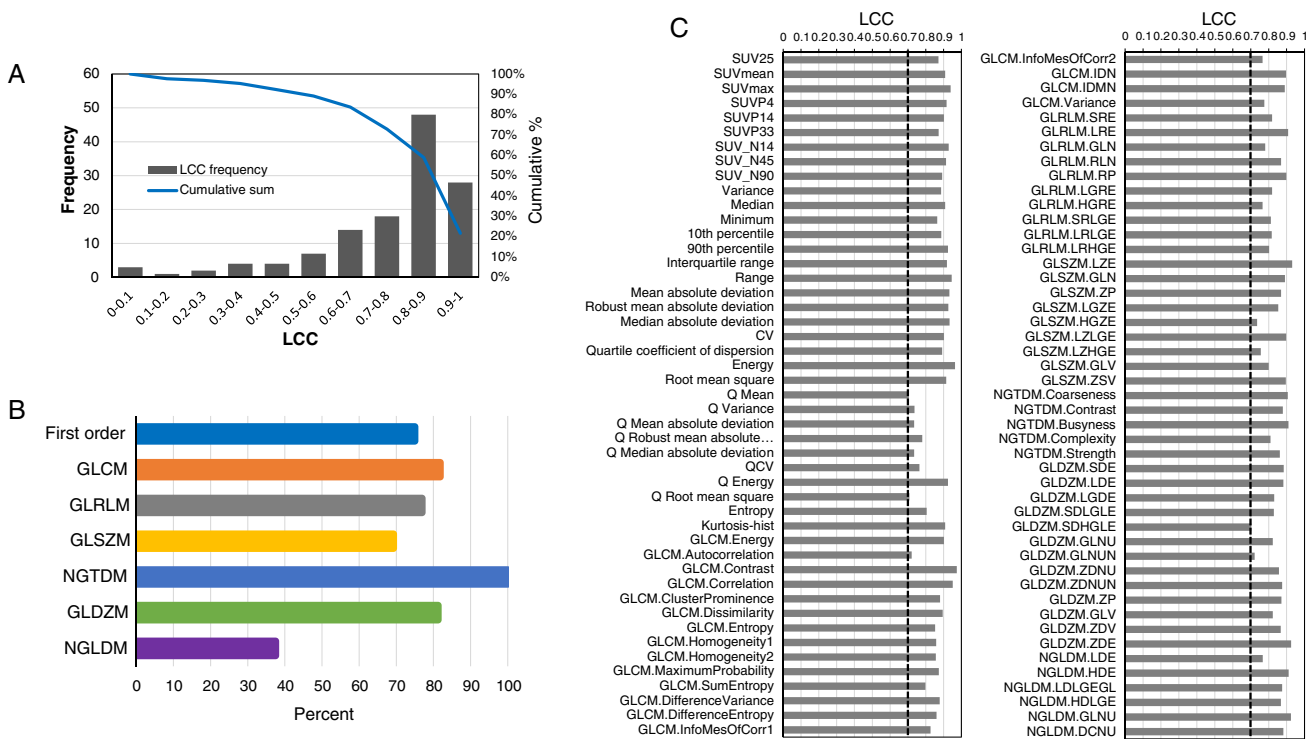


Fig. 3 Repeatability of preclinical radiomic features. **A** Frequency (bar plot, left Y-axis) of repeatability by LCC values and cumulative percent (solid line, right Y-axis). **B** Percent of repeatable features

features. Supplemental Table S1 summarizes the reproducibility of all 131 features.

Cross-correlation between features (preclinical and clinical)

We ascertained the cross-correlation between features using the Spearman correlation (ρ). Highly correlated features ($\rho \geq 0.9$) were removed and reduced to 94 features from 129 features. Hierarchical clustering of the Spearman correlation heatmap is shown Fig. 4. Twenty-one clusters were identified in the preclinical heatmap (Fig. 4A) and similarly 21 clusters were identified in the clinical cross-correlation heatmap (Fig. 4B). Membership of features to clusters is available in Supplemental Table S2. The distribution of Spearman correlations is available in Fig. 4C and D for preclinical and clinical cross-correlations, respectively.

Volume-dependent radiomic features (preclinical and clinical)

In total, 10 radiomic features were highly correlated to volume ($\rho > 0.9$; $P < 0.001$). The functional form of the volume dependency and corresponding goodness-of-fit measures for preclinical and corresponding clinical images is shown in Fig. 5, which was similar for both preclinical and clinical

with $LCC \geq 0.7$ by class (left Y-axis labels). **C** Reproducible radiomic features with $LCC \geq 0.7$ (denoted by dashed vertical black line)

features. Supplemental Table S3 summarizes the statistical analyses for the correlations.

Prediction and assessment of response to therapy

At the intersection of robustness analyses, 62 of the 129 (48.06%) features were found to be optimal and were passed to ReliefF feature selection followed by ML. ReliefF rank importance identified top performing 15 features for prediction (based on BL features) and assessment (based on 4D-BL features) of response to therapy (Fig. 6B and C, respectively). The rank importance of radiomic features is given in Supplemental Table S4.

Preclinical PDX studies

The accuracy of ML in predicting/assessing response to therapy as a function of the number of radiomic features is depicted in Fig. 6D and E for BL and 4D-BL, respectively. The number of radiomic features to maximize prediction accuracy saturated at 4 features (Fig. 6D) with NB exhibiting the highest accuracy at 86.21%, followed by SVM and CART. In contrast, the accuracy of assessing response to therapy (4D-BL) increased with increasing number of radiomic features; the accuracy of NB is 86.9% followed by SVM and CART (Fig. 6E). We opted to

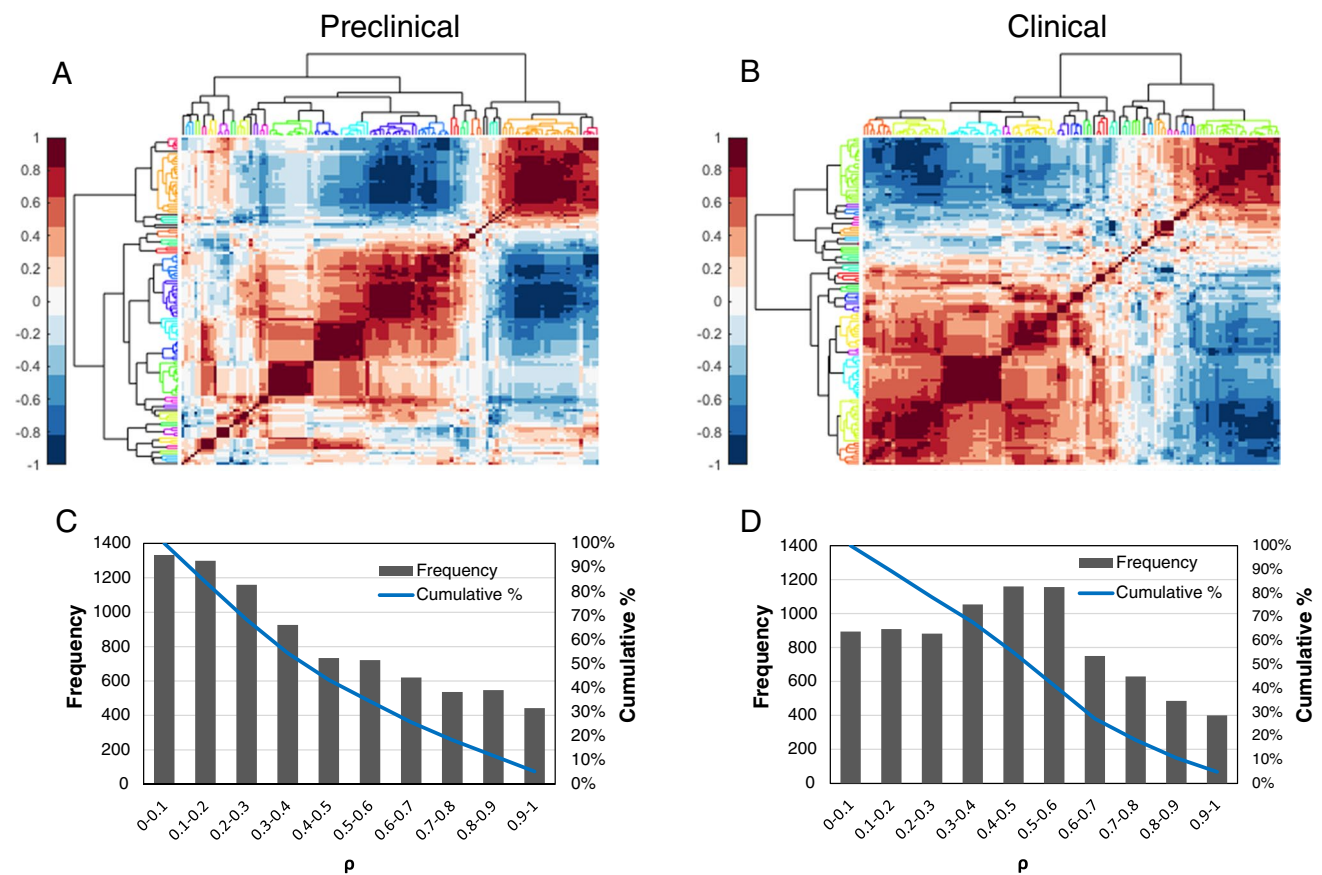


Fig. 4 Cross-correlation between radiomic features. Hierarchical clustering of the Spearman cross-correlation heatmap for **A** preclinical and **B** clinical. The dendrogram shows the clustering, and each

color represents a different cluster. The frequency and cumulative sum at each Spearman correlation (ρ) is displayed in **C** for preclinical and **D** for clinical

compare performance between prediction and assessment (i.e., BL vs. 4D-BL) using the least number of robust features. For this reason, Table 1 tabulates the performance of ML algorithms to predict/assess response prior to and following optimization for robust features using only the top 4 radiomic features for each classification (prediction vs. assessment of response). The set of 4 radiomic features from each task (prediction and assessment of response) make up the radiomics signature (RadSig). As tabulated in Table 1, RadSig performs as well as, or marginally better than, non-optimized features (all features) in predicting response. The performance of prediction/assessment of response to therapy stratified by TNBC subtype is tabulated in Supplemental Table S5 and highlights differences in prediction by TNBC subtype.

The performance of RadSig in comparison to SUV_{mean} , SUV_{max} , and SUL_{peak} for the top two performing ML algorithms (NB and SVM) is summarized in Fig. 7. NB performed marginally better than SVM in predicting/assessing response to therapy (Fig. 7A) in the preclinical PDX trial. The percent increase in predicting/assessing response

to therapy relative to SUV_{mean} , SUV_{max} , and SUL_{peak} is depicted in Fig. 7B for NB. NB-RadSig improved prediction of response by over 60% in all performance measures. In assessing response to therapy, RadSig performed better than SUV_{mean} in most performance criteria and marginally better than SUL_{peak} and SUV_{max} (Fig. 7B). Thus, RadSig has greater impact in predicting response to therapy than assessing response to therapy. Full performance data is available in Supplemental Table S6. We then performed an interim analysis of the ongoing clinical trial to assess the feasibility of implementing PDX-optimized RadSig to predict/assess response to therapy using ML.

Table 2 summarizes patient characteristics, pathological response, SUV metrics at BL, and percent change in SUV metrics between on-treatment (post C1) and BL for the interim analyses (Supplemental Table S7 contains SUV values at baseline and on-treatment). Of the twenty patients, ten patients exhibited pCR; however, all patients exhibited reduction in SUV. Average percent (\pm 1SD) reduction in the non-pCR group was -46.94 ± 21.56 , -53.20 ± 19.91 , and -51.33 ± 19.78 for SUV_{mean} , SUL_{peak} , and SUV_{max} , respectively, and -57.70 ± 14 .

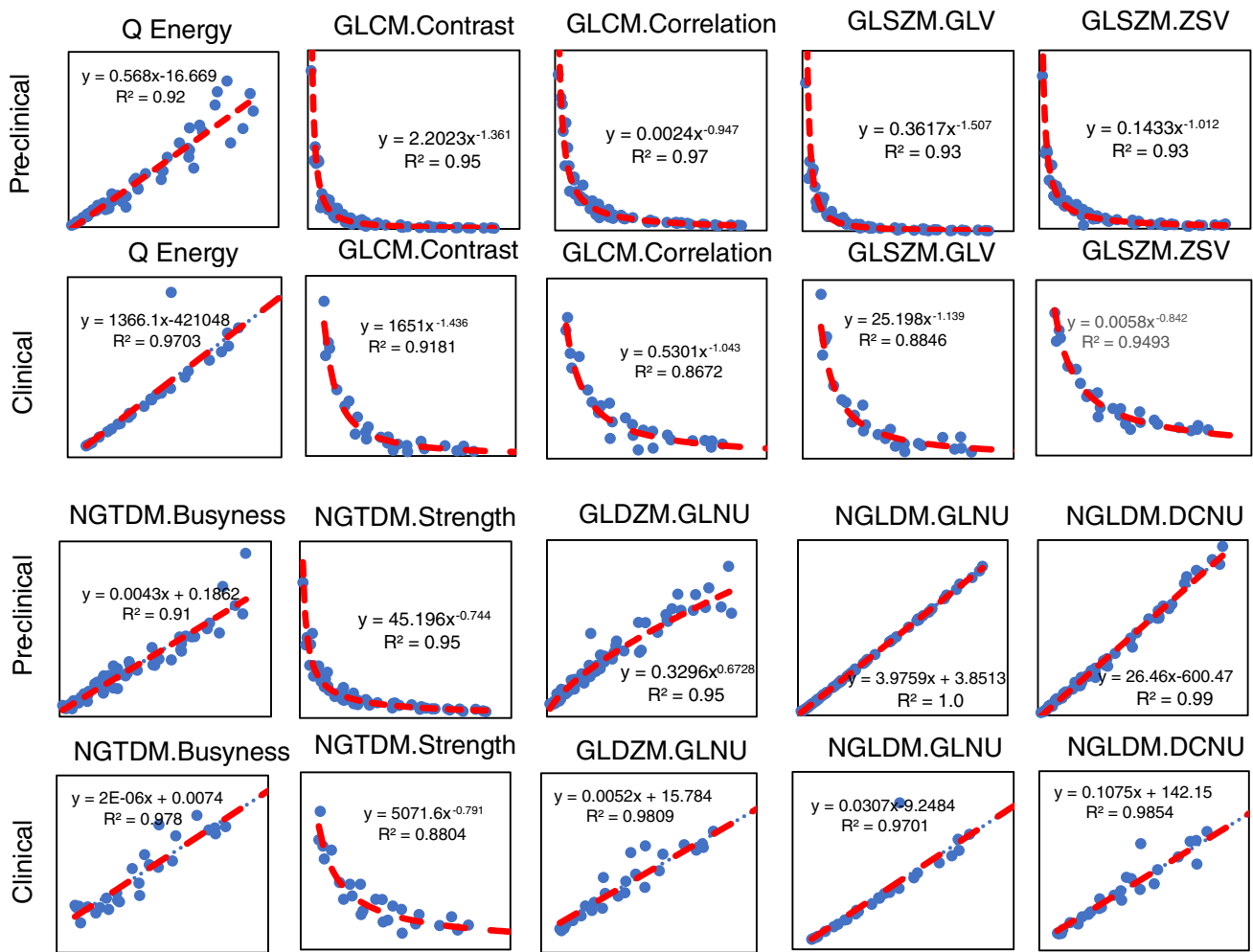


Fig. 5 Volume-dependent preclinical and clinical radiomic features. The correlation between features and tumor volume was assessed for pre-clinical and clinical segmented tumors. Ten common preclinical and clinical features exhibited high correlation ($\rho \geq 0.9$) with tumor volume

83, -60.32 ± 16.47 , and -66.16 ± 13.74 for SUV_{mean} , SUL_{peak} , and SUV_{max} , respectively, in the pCR group. Figure 7 also depicts the performance of the ML algorithms in predicting and assessing response to therapy in the clinical arm (Fig. 7C). The performance of SVM and NB with RadSig as a predictor was marginally similar, although overall SVM performed better than NB when using SUV metrics as predictors (Supplemental Table S6). SVM-RadSig exhibited higher prediction rates of response to therapy relative to SUV_{max} , SUV_{mean} , and SUL_{peak} in all performance measures (20–40% higher), as well as in assessing response to therapy (15–75% higher) (Fig. 7D). Overall, RadSig performed better than SUV metrics in predicting and assessing response to therapy.

Discussion

The emergence of co-clinical models is largely motivated by the realization that established cell lines do not recapitulate the heterogeneity of human tumors and the diversity of tumor phenotypes [11] and that better oncology models are needed to support high-impact translational cancer research [12, 16, 21]. An underlying premise in the co-clinical study design is that the heterogeneity of the human tumor is retained in PDX. Indeed, tumor genomic and pathological investigations have confirmed that PDX recapitulate the heterogeneity of human tumors [12–16]

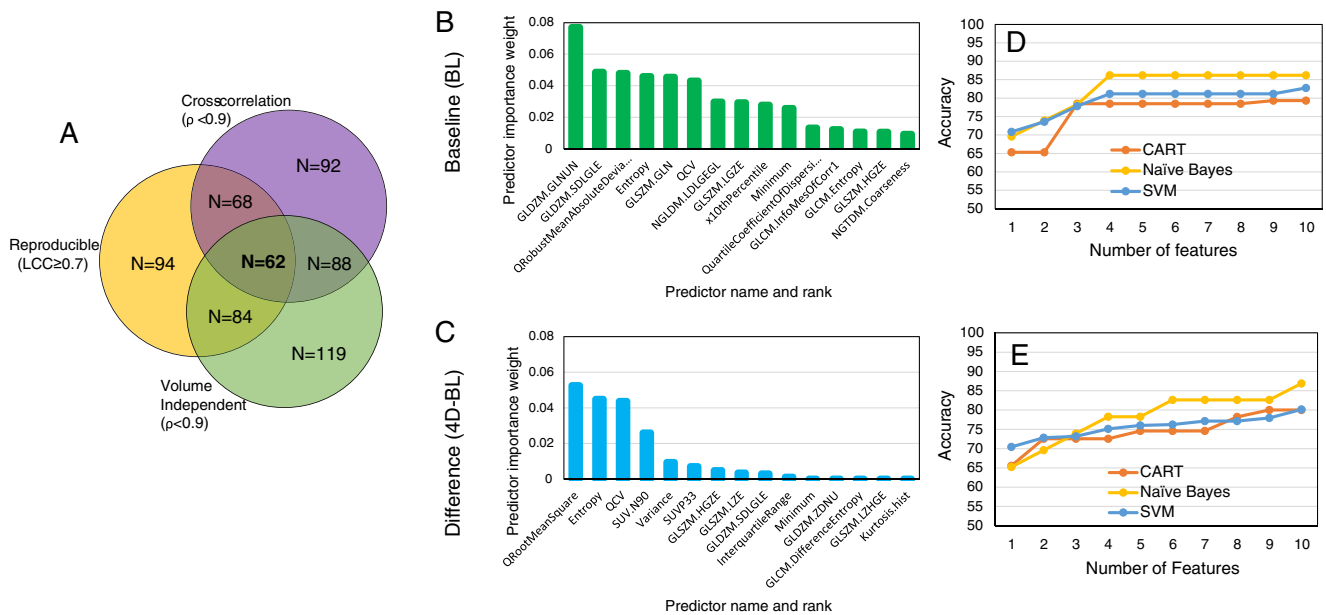


Fig. 6 ML-based selection of radiomic features. **A** At the thresholds defined within to screen for volume dependency, repeatability, and cross-correlation, we identified 62 optimal features. Implementation of relief-based algorithm (RBA) to identify a subset of features as

inputs to ML-based prediction (**B**) and assessment (**C**) of response to therapy. Accuracy of ML algorithms CART, SVM, and Naïve Bayes to predict (**D**) or assess (**E**) response to therapy as a function of number of radiomic features in PDX

Table 1 Accuracy of predicting (BL) and assessing (4D-BL) response to therapy using top 4 radiomic features

Methods	All features		RadSig	
	Prediction	Assessment	Prediction	Assessment
CART	80.34	74.86	78.48	72.57
Naïve Bayes	82.62	82.76	86.21	78.26
SVM	78.48	78.45	81.14	75.13

and that these can be used to better inform cancer biology, therapeutic design [17–19], and therefore by extension imaging studies, albeit with some limitations [21]. With that in mind, in this this work, we exploited the heterogeneity of TNBC PDX subtypes to (1) identify robust radiomic features in preclinical TNBC PDX; (2) optimize RadSig-ML algorithms to predict response to therapy in PDX; and (3) implement PDX-optimized RadSig to predict/assess response to therapy in the clinical trial.

To our knowledge, this study represents the first such effort to optimize radiomic features in preclinical PET imaging to predict/assess response to therapy in TNBC PDX. We recently characterized the dependency of preclinical MR radiomic features on tumor volume [37]. In this work, we confirmed dependency of preclinical PET radiomic features on tumor volume with strikingly similar clinical parallels. This is particularly relevant in longitudinal studies during which tumor volumes will change with the course of the

disease or following therapy. Ideally, volume-independent features should be used as to not bias image features longitudinally. We further evaluated the cross-correlation of pre-clinical and clinical radiomic features with the goal of reducing the dimensionality of features. Finally, we evaluated the repeatability of radiomic features in preclinical PET imaging to identify robust features for inclusion in ML-based prediction of response to therapy. At the thresholds defined within to screen for volume dependency, repeatability, and cross-correlation, we identified 62 optimal features to predict/assess response to therapy.

RBF [27] was used to rank image features using three ML algorithms as to their relevance in predicting/assessing response to therapy. Our data suggests that overall SVM performed better than NB and CART in predicting response to therapy. We used the top four ML-RBF-optimized radiomic features—referred to as radiomic signature (RadSig)—from each task (prediction vs. assessment) to either predict or assess response to therapy. In the preclinical arm, RadSig performed significantly better in predicting response to therapy relative to standard SUV measures. Importantly, RadSig also performed better in predicting and assessing response to therapy in the clinical arm. Antunovic et al. [38] reported the utility of FDG-PET radiomic features to assess response to therapy using four different models in 79 patients with heterogenous breast cancer subtypes. The reported area under the curve of an ROC analysis ranged from 0.70 to 0.73. Li et al. [39] recently assessed the utility of both PET

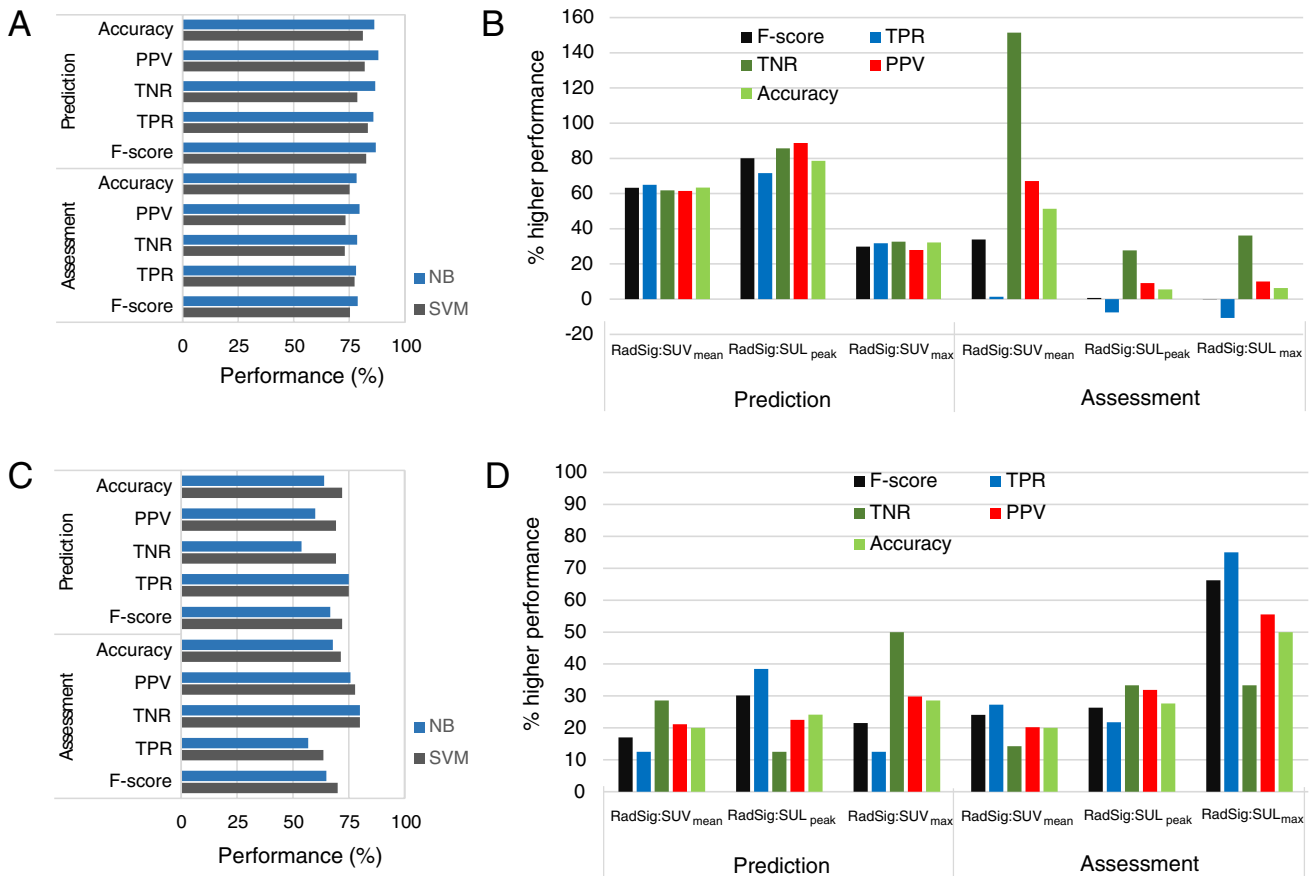


Fig. 7 Performance of ML algorithms. **A** Performance of RadSig in predicting/assessing response to therapy in the preclinical PDX trial with NB and SVM. **B** Percent improvement in NB-RadSig prediction/

assessment of response relative to SUV_{max}, SUV_{mean}, and SUL_{peak}. **C** similar to **A** but for the clinical investigation. **D** Similar to **B** but using SVM

and CT radiomic features to predict response to therapy in a retrospective study that included 100 heterogenous breast cancer patients. The PET/CT radiomic predictors achieved a prediction accuracy of 87% on the training split set and 77% on the independent validation set.

In this small, albeit homogenous, dataset of TNBC patients where PDX-optimized radiomic features were implemented in the clinical imaging arm, we observed an impressive accuracy of 72% and 71% when predicting and assessing response, respectively, compared to SUV metrics. We were unable to perform a validation test on an independent dataset. However, the primary objective was to compare the performance of predictive metrics in the training phase relative to standard SUV measures. Prediction of response to therapy can be further enhanced through integration of MR image features [40]. At this stage, we did not include MR radiomic features due to increased dimensionality with added MR image features and a limited number of patients. Other technologies that could be integrated with imaging to enhance therapeutic prediction include liquid biopsies such as circulating tumor DNA (ctDNA) analyses [41] and molecular/genomic features of

tumors [42], both of which are an active area of investigation. Finally, numerous recent studies have documented that pCR rates varied with breast cancer molecular subtypes. TNBC and HER2-positive molecular subtypes have shown to have higher pCR rates after NAC [43]. Importantly, several studies have demonstrated an association between imaging features and molecular phenotypes, risk of recurrence, and prognosis [44–46]. Interestingly, our PDX studies similarly suggest that response to therapy (and prediction thereof) is a function of the TNBC subtype; however, further studies are needed to support this hypothesis and the utility of radiomic features in classifying TNBC subtypes. With that in mind, one of the most critical aspects in practical implementation of radiomics is a consensus on the most effective features and their standardization.

Conclusions

We identified robust FDG-PET radiomic features in terms of volume dependency, reproducibility, and cross-correlation to predict and assess response to therapy in a

Table 2 Patient characteristics, pathologic response, and SUV metrics

Stage at diagnosis	Grade at diag	pCR	BL SUV _{mean}	BL SUL _p	BL SUV _{max}	%Δ SUV _{mean}	%Δ SUL _{peak}	%Δ SUV _{max}
IIB (T2N1)	2	No	1.86	1.57	4.03	NA	NA	NA
IIB (T2N1)	3	Yes	6.59	4.58	11.27	−60.97	−68.11	−57.82
IIIA (T3N1)	3	No	10.55	12.34	26.41	NA	NA	NA
IIA (T2N0)	?	Yes	8.77	7.80	21.48	−75.49	−82.21	−86.77
IIB (T2N1)	3	No	3.18	1.64	5.31	−4.32	−13.19	−9.40
IIB (T2N1)	3	No	2.92	1.29	4.54	−38.67	−41.17	−41.77
			8.01	6.43	15.55	−40.82	−44.37	−42.90
IIB (T2N1)	3	No	8.58	6.20	17.02	−64.92	−70.33	−73.40
			8.91	6.62	20.11	−62.72	−68.17	−61.72
IIA (T2N0)	3	Yes	2.49	1.61	5.25	NA	NA	NA
IIA (T2N0)	3	Yes	4.17	3.06	12.45	−48.84	−48.53	−64.23
IIA (T2N0)	3	Yes	3.89	3.12	8.62	−57.76	−59.91	−73.72
IIB (T2N1)	3	No	2.85	2.59	6.28	−41.45	−55.68	−55.97
			2.48	2.03	5.03	−27.04	−37.81	−49.88
IIA (T2N0)	3	No	8.66	7.56	21.87	−75.82	−82.48	−81.35
IIB (T2N1)	3	Yes	10.29	8.32	28.33	−74.14	−76.39	−81.07
			3.64	2.02	5.65	−58.93	−57.11	−55.17
			8.00	5.60	16.06	−71.75	−77.22	−77.14
IIA (T2N0)	3	Yes	1.86	1.65	4.18	−30.42	−36.91	−53.02
IIA (T2N0)	3	No	9.80	7.74	17.69	−67.65	−64.44	−54.15
IIA (T2N0)	3	Yes	7.19	5.21	14.57	−46.36	−42.49	−43.23
IIA (T2N0)	3	No	7.98	7.26	18.46	−45.99	−54.32	−42.74
IIA (T2N0)	3	Yes	2.91	2.03	5.24	−40.23	−40.04	−58.66
IIB (T2N1)	3	Yes	6.19	5.53	11.28	−69.81	−74.57	−76.91
IIB (T2N1)	3	No	2.44	1.99	5.53	NA	NA	NA

preclinical PDX trial. The number of radiomic features to maximize accuracy was further optimized to yield ML radiomic signatures (RadSig) of response to therapy. RadSig improved prediction of response to therapy in the preclinical arm. Given that PDX recapitulate the heterogeneity of human tumors, we then assessed the feasibility of implementing PDX-optimized RadSig in an interim analysis of the clinical trial to predict response to therapy. The performance of SVM-RadSig in predicting/assessing response to therapy was superior to SUV_{max}, SUV_{mean}, and SUL_{peak} metrics in the clinical setting; however, given the small sample size, additional studies are warranted to further validate the utility of PDX-optimized features, such as RadSig, and potentially integrate with multi-scale features to enhance prediction/assessment of response to therapy.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00259-021-05489-8>.

Acknowledgements The authors acknowledge the staff of the Preclinical Imaging Facility and the Center for Clinical Imaging Research (CCIR) at Mallinckrodt Institute of Radiology (MIR), Washington University School of Medicine, for performing imaging studies.

Author contribution Conceptualization: SR, FOA, and KIS; methodology: SR, TDW, SL, and KIS; formal analysis and investigation: SR and KIS; writing—original draft preparation: SR; writing—review and editing: RLW, FD, and KIS; funding acquisition: RWL, FOA, SL, and KIS; resources: SL; supervision: FD and KIS. All authors read and approved the final manuscript.

Funding This work was supported by NCI grants U24CA209837, U24CA253531, and U54CA224083; U2CCA233303, and K12CA167540; Siteman Cancer Center (SCC) Support Grant P30CA091842; and Internal funds provided by Mallinckrodt Institute of Radiology.

Availability of data and material All the co-clinical data will be available for download through the Washington University School of Medicine Co-Clinical Imaging Research Resource web portal at <https://c2ir2.wustl.edu/>, co-clinical database (CCDB).

Code availability Not applicable.

Declarations

Ethics approval All studies were performed with approval from the Washington University Humans subjects research committee and animal studies committee.

Consent to participate Informed consent to participate in the study was obtained from all participants.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

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/>.

References

1. Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the rosetta stone of therapy resistance. *Cancer Cell*. 2020;37:471–84. <https://doi.org/10.1016/j.ccell.2020.03.007>.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278:563–77. <https://doi.org/10.1148/radiol.2015151169>.
3. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. <https://doi.org/10.1038/ncomms5006>.
4. Chen Z, Akbay E, Mikse O, Tupper T, Cheng K, Wang Y, et al. Co-clinical trials demonstrate superiority of crizotinib to chemotherapy in ALK-rearranged non-small cell lung cancer and predict strategies to overcome resistance. *Clin Cancer Res*. 2014;20:1204–11. <https://doi.org/10.1158/1078-0432.Ccr-13-1733>.
5. Kim HR, Kang HN, Shim HS, Kim EY, Kim J, Kim DJ, et al. Co-clinical trials demonstrate predictive biomarkers for dovitinib, an FGFR inhibitor, in lung squamous cell carcinoma. *Ann Oncol*. 2017;28:1250–9. <https://doi.org/10.1093/annonc/mdx098>.
6. Kwong LN, Boland GM, Frederick DT, Helms TL, Akid AT, Miller JP, et al. Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. *J Clin Invest*. 2015;125:1459–70. <https://doi.org/10.1172/jci78954>.
7. Lunardi A, Ala U, Epping MT, Salmena L, Clohessy JG, Webster KA, et al. A co-clinical approach identifies mechanisms and potential therapies for androgen deprivation resistance in prostate cancer. *Nat Genet*. 2013;45:747–55. <https://doi.org/10.1038/ng.2650>.
8. Nishino M, Sacher AG, Gandhi L, Chen Z, Akbay E, Fedorov A, et al. Co-clinical quantitative tumor volume imaging in ALK-rearranged NSCLC treated with crizotinib. *Eur J Radiol*. 2017;88:15–20. <https://doi.org/10.1016/j.ejrad.2016.12.028>.
9. Owonikoko TK, Zhang G, Kim HS, Stinson RM, Bechara R, Zhang C, et al. Patient-derived xenografts faithfully replicated clinical outcome in a phase II co-clinical trial of arsenic trioxide in relapsed small cell lung cancer. *J Transl Med*. 2016;14:111. <https://doi.org/10.1186/s12967-016-0861-5>.
10. Sia D, Moeini A, Labgaa I, Villanueva A. The future of patient-derived tumor xenografts in cancer treatment. *Pharmacogenomics*. 2015;16:1671–83. <https://doi.org/10.2217/pgs.15.102>.
11. Sulaiman A, Wang L. Bridging the divide: preclinical research discrepancies between triple-negative breast cancer cell lines and patient tumors. *Oncotarget*. 2017;8:113269–81. <https://doi.org/10.18632/oncotarget.22916>.
12. DeRose YS, Wang G, Lin YC, Bernard PS, Buys SS, Ebbert MT, et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat Med*. 2011;17:1514–20.
13. Zhao X, Liu Z, Yu L, Zhang Y, Baxter P, Voicu H, et al. Global gene expression profiling confirms the molecular fidelity of primary tumor-based orthotopic xenograft mouse models of medulloblastoma. *Neuro-Oncol*. 2012;14:574–83.
14. Morton CL, Houghton PJ. Establishment of human tumor xenografts in immunodeficient mice. *Nat Protoc*. 2007;2:247–50. <https://doi.org/10.1038/nprot.2007.25>.
15. Reyat F, Guyader C, Decraene C, Lucchesi C, Auger N, Assayag F, et al. Molecular profiling of patient-derived breast cancer xenografts. *Breast Cancer Res*. 2012;14:R11.
16. Krepler C, Xiao M, Spoesser K, Brafford PA, Shannan B, Beqiri M, et al. Personalized pre-clinical trials in BRAF inhibitor resistant patient derived xenograft models identify second line combination therapies. *Clin Cancer Res*. 2015. <https://doi.org/10.1158/1078-0432.CCR-15-1762>.
17. Kerbel RS. Human tumor xenografts as predictive preclinical models for anticancer drug activity in humans: better than commonly perceived-but they can be improved. *Cancer Biol Ther*. 2003;2:S134–9.
18. Johnson JI, Decker S, Zaharevitz D, Rubinstein LV, Venditti JM, Schepartz S, et al. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *Br J Cancer*. 2001;84:1424–31. <https://doi.org/10.1054/bjoc.2001.1796>.
19. Scholz CC, Berger DP, Winterhalter BR, Henss H, Fiebig HH. Correlation of drug response in patients and in the clonogenic assay with solid human tumour xenografts. *Eur J Cancer*. 1990;26:901–5.
20. Savaikar MA, Whitehead T, Roy S, Strong L, Fettig N, Prmeau T, et al. Preclinical PERCIST and 25% of SUVmax threshold: precision imaging of response to therapy in co-clinical (18)F-FDG PET imaging of triple-negative breast cancer patient-derived tumor xenografts. *J Nucl Med*. 2020;61:842–9. <https://doi.org/10.2967/jnumed.119.234286>.
21. Shoghi KI, Badea CT, Blocker SJ, Chenevert TL, Laforest R, Lewis MT, et al. Co-clinical imaging resource program (CIRP): bridging the translational divide to advance precision medicine. *Tomography*. 2020;6:273–87. <https://doi.org/10.18383/j.tom.2020.00023>.
22. Li S, Shen D, Shao J, Crowder R, Liu W, Prat A, et al. Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep*. 2013;4:1116–30. <https://doi.org/10.1016/j.celrep.2013.08.022>.
23. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121:2750–67. <https://doi.org/10.1172/JCI45014>.
24. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009;50(Suppl 1):122S-S150. <https://doi.org/10.2967/jnumed.108.057307>.
25. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput

- image-based phenotyping. *Radiology*. 2020;295:328–38. <https://doi.org/10.1148/radiol.2020191145>.
26. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE T Syst Man Cyb*. 1973;Smc3:610–21. <https://doi.org/10.1109/Tsmc.1973.4309314>.
 27. Robnik-Sikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn*. 2003;53:23–69. <https://doi.org/10.1023/A:1025667309714>.
 28. Urbanowicz RJ, Meeker M, La Cava W, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. *J Biomed Inform*. 2018;85:189–203. <https://doi.org/10.1016/j.jbi.2018.07.014>.
 29. Cheng Z, Nakatsugawa M, Hu C, Robertson SP, Hui X, Moore JA, et al. Evaluation of classification and regression tree (CART) model in weight loss prediction following head and neck cancer radiation therapy. *Adv Radiat Oncol*. 2018;3:346–55. <https://doi.org/10.1016/j.adro.2017.11.006>.
 30. Huang MW, Chen CW, Lin WC, Ke SW, Tsai CF. SVM and SVM ensembles in breast cancer prediction. *Plos One*. 2017;12. <https://doi.org/10.1371/journal.pone.0161501>.
 31. Gao HY, Zeng X, Yao CH. Application of improved distributed naive Bayesian algorithms in text classification. *J Supercomput*. 2019;75:5831–47. <https://doi.org/10.1007/s11227-019-02862-1>.
 32. Murali N, Kucukkaya A, Petukhova A, Onofrey J, Chapiro J. Supervised machine learning in oncology: a clinician's guide. *Dig Dis Interv*. 2020;4:73–81. <https://doi.org/10.1055/s-0040-1705097>.
 33. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in oncological PET/CT: a methodological overview. *Nucl Med Mol Imaging*. 2019;53:14–29. <https://doi.org/10.1007/s13139-019-00571-4>.
 34. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45:255–68.
 35. Groutz A, Blaivas JG, Chaikin DC, Resnick NM, Engleman K, Anzalone D, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: a multicenter study of micturition diary and pad tests. *J Urol*. 2000;164:698–701. <https://doi.org/10.1097/00005392-200009010-00019>.
 36. Matheson GJ. We need to talk about reliability: making better use of test-retest studies for study design and interpretation. *PeerJ*. 2019;7:e6918. <https://doi.org/10.7717/peerj.6918>.
 37. Roy S, Whitehead TD, Quirk JD, Salter A, Ademuyiwa FO, Li S, et al. Optimal co-clinical radiomics: sensitivity of radiomic features to tumour volume, image noise and resolution in co-clinical T1-weighted and T2-weighted magnetic resonance imaging. *EBioMedicine*. 2020;59:102963. <https://doi.org/10.1016/j.ebiom.2020.102963>.
 38. Antunovic L, De Sanctis R, Cozzi L, Kirienko M, Sagona A, Torrisi R, et al. PET/CT radiomics in breast cancer: promising tool for prediction of pathological response to neoadjuvant chemotherapy. *Eur J Nucl Med Mol Imaging*. 2019;46:1468–77. <https://doi.org/10.1007/s00259-019-04313-8>.
 39. Li P, Wang X, Xu C, Liu C, Zheng C, Fulham MJ, et al. (18) F-FDG PET/CT radiomic predictors of pathologic complete response (pCR) to neoadjuvant chemotherapy in breast cancer patients. *Eur J Nucl Med Mol Imaging*. 2020;47:1116–26. <https://doi.org/10.1007/s00259-020-04684-3>.
 40. Hu Q, Whitney HM, Giger ML. Radiomics methodology for breast cancer diagnosis using multiparametric magnetic resonance imaging. *J Med Imaging (Bellingham)*. 2020;7:044502. <https://doi.org/10.1117/1.JMI.7.4.044502>.
 41. Magbanua MJM, Swigart LB, Wu HT, Hirst GL, Yau C, Wolf DM, et al. Circulating tumor DNA in neoadjuvant-treated breast cancer reflects response and survival. *Ann Oncol*. 2021;32:229–39. <https://doi.org/10.1016/j.annonc.2020.11.007>.
 42. Zanfardino M, Franzese M, Pane K, Cavaliere C, Monti S, Esposito G, et al. Bringing radiomics into a multi-omics framework for a comprehensive genotype-phenotype characterization of oncological diseases. *J Transl Med*. 2019;17:337. <https://doi.org/10.1186/s12967-019-2073-2>.
 43. Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2018;170:559–67. <https://doi.org/10.1007/s10549-018-4801-3>.
 44. Li W, Yu K, Feng C, Zhao D. Molecular subtypes recognition of breast cancer in dynamic contrast-enhanced breast magnetic resonance imaging phenotypes from radiomics data. *Comput Math Methods Med*. 2019;2019:6978650. <https://doi.org/10.1155/2019/6978650>.
 45. Wang Q, Mao N, Liu M, Shi Y, Ma H, Dong J, et al. Radiomic analysis on magnetic resonance diffusion weighted image in distinguishing triple-negative breast cancer from other subtypes: a feasibility study. *Clin Imaging*. 2021;72:136–41. <https://doi.org/10.1016/j.clinimag.2020.11.024>.
 46. Huang SY, Franc BL, Harnish RJ, Liu G, Mitra D, Copeland TP, et al. Exploration of PET and MRI radiomic features for decoding breast cancer phenotypes and prognosis. *NPJ Breast Cancer*. 2018;4:24. <https://doi.org/10.1038/s41523-018-0078-2>.

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